UNIVERSITÉ DE BORDEAUX II

FACULTÉ DE PHARMACIE

1971 — N° 628

ÉTUDE DU RAPPORT DE DEUX VARIABLES ALÉATOIRES NORMALES. APPLICATIONS A LA BIOLOGIE

THÈSE POUR LE DOCTORAT DE L'UNIVERSITÉ

Mention « PHARMACIE »

PAR

MICHEL AHYERRE

PHARMACIEN-CHIMISTE DES ARMÉES

DIPLÔMÉ DES CERTIFICATS D'ÉTUDES SUPÉRIEURES

DE CHIMIE ORGANIQUE, CHIMIE PHYSIQUE ET MICROBIOLOGIE

ATTACHÉ DE FACULTÉ, ASSISTANT DE SCIENCES FONDAMENTALES AU LABORATOIRE DE PHYSIQUE MÉDICALE

né à Saint-Dizier (Haute-Marne) le 7 juin 1948

Présentée et soutenue publiquement le jeudi 28 octobre 1971

Examinateurs de la thèse :

MM. P. BLANQUET, Professeur ..... *Président*

J. DUHAMEL, Professeur .....

C. VRIGNAUD, Professeur ..... *Juges*

D. DUCASSOU, Maître de Conférences agrégé .....

1971

Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41543508_0012

UNIVERSITÉ DE BORDEAUX II

FACULTÉ DE PHARMACIE

1971 — N° 628

ÉTUDE DU RAPPORT DE DEUX VARIABLES ALÉATOIRES NORMALES. APPLICATIONS A LA BIOLOGIE

THÈSE POUR LE DOCTORAT DE L'UNIVERSITÉ

Mention « PHARMACIE »

PAR

MICHEL AHYERRE

PHARMACIEN-CHIMISTE DES ARMÉES

DIPLÔMÉ DES CERTIFICATS D'ÉTUDES SUPÉRIEURES

DE CHIMIE ORGANIQUE, CHIMIE PHYSIQUE ET MICROBIOLOGIE

ATTACHÉ DE FACULTÉ, ASSISTANT DE SCIENCES FONDAMENTALES AU LABORATOIRE DE PHYSIQUE MÉDICALE

né à Saint-Dizier (Haute-Marne) le 7 juin 1948

Présentée et soutenue publiquement le jeudi 28 octobre 1971

Examinateurs de la thèse :

MM. P. BLANQUET, Professeur ..... *Président*

J. DUHAMEL, Professeur ..... *Juges*

C. VRIGNAUD, Professeur ..... *Juges*

D. DUCASSOU, Maître de Conférences agrégé ..... *Juges*

1971

UNIV.
LEIDEN
BIBL.

A MES PARENTS

avec toute l'affection que je leur porte

A MA FEMME

en témoignage de mon amour.

A MES FRERES ET BEAUX-FRERES

A MES BEAUX-PARENTS

A TOUTE MA FAMILLE.

A MES CAMARADES DU LABORATOIRE

A MADAME POMMIER

pour son amabilité et son dévouement.

A MES AMIS DE L'ECOLE PRINCIPALE DU SERVICE DE SANTE DE LA MARINE.

A MONSIEUR LE PROFESSEUR AGREGE J.C. ROQUES,

Vous nous avez toujours conseillé avec une très grande gentillesse.

Nous avons largement abusé de votre aide, qui, au cours de l'élaboration de ce travail nous a été très précieuse.

Veuillez trouver ici l'expression de nos vifs remerciements

A MONSIEUR LE MEDECIN GENERAL DE 2ème CLASSE, CURET

Spécialiste des Hôpitaux des Armées

Chevalier de la Légion d'Honneur

Officier de l'Ordre National du Mérite

Croix de guerre 1939-1945

Directeur de l'Ecole Principale du Service de Santé de la Marine.

A MONSIEUR LE MEDECIN EN CHEF DE 1ère CLASSE, LESTAGE

Chevalier de la Légion d'Honneur

Sous-directeur de l'Ecole Principale du service de Santé de la Marine.

A nos Juges,

Monsieur LE PROFESSEUR J. DUHAMEL

Professeur de Physique Médicale et Pharmaceutique
Ancien Elève de l'Ecole Polytechnique
Officier des Palmes Académiques
Chevalier de la Santé publique
Electroradiologiste des Hôpitaux

Vous nous avez proposé le sujet de cette thèse, et, pas à pas vous en avez suivi le déroulement. Vous avez consacré un temps précieux à nous initier au calcul statistique.

Toujours prêt à donner un conseil, vous avez guidé nos travaux avec une exceptionnelle bienveillance.

Nous vous prions, Monsieur le Professeur, de bien vouloir agréer l'expression de notre immense gratitude.

Monsieur le Professeur C. VRIGNAUD

Professeur sans chaire (Physique pharmaceutique)

Nous avons bénéficié de votre enseignement au cours de nos études. Vous nous faites l'honneur de siéger dans notre jury de thèse. Soyez assuré, Monsieur le Professeur, de notre profonde reconnaissance.

Monsieur le Docteur D. DUCASSOU

Maître de Conférences agrégé
Biologiste des Hôpitaux
Licencié és-Sciences

Vous avez bien voulu accepter de juger notre travail. Veuillez trouver ici le témoignage de nos sentiments de vive gratitude.

A notre Président de thèse,

MONSIEUR LE PROFESSEUR P. BLANQUET

Professeur de Physique nucléaire appliquée
et de radiobiologie.
Docteur és-sciences-Physiques
Officier des Palmes Académiques
Médaille de Vermeil du Service de Santé de la Marine
Visiting professor de l'Université de Californie
Berkeley - U.S.A.

Nous vous prions de bien vouloir trouver ici, pour l'honneur que vous nous faites en présidant le jury de cette thèse, l'expression de notre reconnaissance et de notre profond respect.

ETUDE DU RAPPORT DE DEUX VARIABLES ALEATOIRES NORMALES

APPLICATION A LA BIOLOGIE

P L A N D ' E T U D E

I - ETUDE THEORIQUE

1 - Propriétés d'une variable aléatoire
Distribution de Laplace-Gauss
Distribution de Cauchy

2 - Fonction d'une seule variable aléatoire
Cas particulier de la transformation linéaire et non linéaire.

3 - Fonction de plusieurs variables aléatoires

4 - Somme de variables aléatoires

5 - Produit de variables aléatoires

6 - Rapport de deux variables aléatoires.

II - ETUDE EXPERIMENTALE

1 - ETUDE DU RAPPORT de deux variables aléatoires normales et liées entre elles: les gamma-globulines et les sérum-albumines.

2 - ETUDE DU RAPPORT de deux variables aléatoires non normales et non liées entre elles:: les variables "EPONGE" et "TOTALITE" intervenant dans le test Tétrasorb.

ETUDE THEORIQUE

I N T R O D U C T I O N

Les données brutes de l'expérience sont très souvent en Biologie ou Médecine, soumises à des calculs ou des anamorphoses, pour ne pas dire triturées.

Souvent même, plusieurs données expérimentales sont combinées dans une fonction en général simple.

Les propriétés de la distribution de ces fonctions de plusieurs variables aléatoires, par leur double caractère d'être entâchées d'erreurs d'expérience et de traduire la variation biologique des individus, sont mal connues. La raison en est que les calculs sont en général ingrats, ou d'une facilité dérisoire, mais peu précis.

Ainsi, nous nous sommes demandés, si les maigres résultats obtenus ne seraient pas vérifiés, dans l'expérience biologique journalière. En outre, nous regrouperons dans ce travail des résultats dont certains sont très importants, et qui sont exposés dans des livres spécialisés.

CHAPITRE PREMIER

PROPRIETES D'UNE VARIABLE ALEATOIRE - DISTRIBUTION DE LAPLACE-GAUSS - DISTRIBUTION DE CAUCHY

I - Propriétés d'une variable aléatoire

Nous ne considérerons que des variables aléatoires continues. (Les résultats sont aussi valables pour des variables aléatoires discontinues).

Soit une variable aléatoire, continue X, pouvant varier entre les valeurs numériques a et b (limites de variation, ces quantités pouvant d'ailleurs être infinies).

Cette variable aléatoire X est définie par sa valeur numérique x.

La probabilité pour que la valeur de X soit comprise entre x et $x + dx$ est $p_X(x)\,dx$.

$p_X(x)$ définit la loi de distribution de X.

On déduit de cette loi de distribution, les résultats suivants :

- La probabilité $p_X(x)$, pour que la valeur de X soit comprise entre a et x, que l'on appelle loi de répartition de X est :

$$(I) \qquad \boxed{P_X(x) = \int_a^x p_X(x)\,dx.}$$

- comme la probabilité que la valeur de X soit comprise entre a et b vaut un (certitude) on peut écrire :

$$(II) \qquad \boxed{\int_a^b p_X(x)\,dx = 1.}$$

- La courbe représentative de la fonction $p_X(x)$ <u>est la courbe de distribution</u> de X.

Nous aurons seulement à considérer des courbes présentant un maximum unique et symétriques par rapport à ce maximum.

Soit μ la valeur de x correspondant à ce maximum.

- Il découle de la symétrie que :

$$(III) \quad \int_a^{\mu} p_X(x)\,dx = \int_{\mu}^{b} p_X(x)\,dx = \frac{1}{2}\,.$$

μ est la valeur médiane de X.

En principe la fonction de distribution peut contenir un nombre quelconque de paramètres. En fait, nous aurons à considérer seulement des lois à deux paramètres.

- l'un de ces paramètres est μ . Il place la courbe de distribution dans l'échelle des abcisses.

- l'autre paramètre est σ : il est lié à la valeur de l'unité qui sert à mesurer la valeur numérique de X.

Nous aurons en outre, à considérer :

1º) L'espérance mathématique de X qui vaut par définition :

$$(IV) \quad E(X) = \int_a^b x\, p_X(x)\,dx$$

2º) Les moments d'ordre k, de X qui se déduisent de la formule précédente en remplaçant l'x en facteur par x^k.

Parmi les moments d'ordre k il en existe un qui est particulier : c'est le moment d'ordre 2.

3º) Les moments centrés d'ordre k de X qui se déduisent en remplaçant dans (IV), la quantité x par $(x - E(X))^k$. Parmi eux, il n'y en a guère qu'un seul de vraiment important dans la pratique courante : le moment d'ordre 2, ou variance. Nous le symboliserons par V(X) :

$$(V) \quad V(X) = \int_a^b \left[x - E(X)\right]^2 p_X(x)\,dx.$$

Dans cette définition :

- l'espérance mathématique de X est prise comme origine

- $x - E(X)$ est la valeur numérique d'une variable aléatoire dont l'espérance mathématique est nulle (on dit que c'est une variable aléatoire centrée).

On peut définir la variance de X comme :

$$V(X) = E \quad (x - E(X))^2$$

Les lois de distribution sont tabulées en supposant que la valeur numérique de la variable aléatoire est une valeur centrée ; supposer que E(X) est différent de 0, revient à une translation d'ordonnée.

D'autre part, elles sont tabulées en supposant que $\sigma = 1$. Supposer une valeur différente revient à affecter un coefficient multiplicatif. Dans ces conditions, on parle de variable normée, et on désigne en général par U une variable aléatoire dont les valeurs numériques centrées, normées sont égales à u.

II - Loi de Laplace-Gauss.

Cette loi jouit d'une réputation flatteuse. Ainsi on s'imagine, à tort ou à raison, que beaucoup de paramètres sont distribués suivant cette loi, et nombre de théorèmes, que nous énoncerons par la suite, lui accordent une position priviligiée.

1º) Formule

$$p(X) = \frac{1}{\sigma\sqrt{2\pi}} \quad \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

2º) Propriétés

- On voit que cette fonction de distribution est symétrique par rapport à la droite $x = \mu$.

- Sa dérivée première est proportionnelle à :

$$(x - \mu) \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Elle s'annule pour $x = \mu$

Ceci montre que la médiane est également le mode.

- La dérivée seconde est proportionnelle à :

$$\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\left(1 - \frac{(x-\mu)^2}{\sigma^2}\right)$$

soit $(x - \mu + \sigma)\ (x - \mu - \sigma)\ \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

On voit que la dérivée seconde s'annule pour $x = \mu - \sigma$ et $x = \mu + \sigma$

Autrement dit, la courbe représentative a deux points d'inflexion équidistants de σ par rapport à la médiane.

3º) Espérance mathématique E (X).

$$E(X) = \int_{-\infty}^{+\infty} \frac{x}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx .$$

puisque la valeur numérique de la variable aléatoire existe pour tout x positif et tout x négatif.

Le résultat est absolument classique lorsqu'on utilise le changement de variable $u = \frac{x - \mu}{\sigma}$

- Dans ces conditions

$dx = \sigma\ du.$

$$E(X) = \int_{-\infty}^{+\infty} \frac{\sigma u + \mu}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du$$

$$E(X) = \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} u \exp\left(-\frac{u^2}{2}\right) du + \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(-\frac{u^2}{2}\right) du$$

La première intégrale vaut :

$$\frac{\sigma}{\sqrt{2\pi}} \exp\left[-\frac{u^2}{2}\right]_{-\infty}^{+\infty} = 0$$

La deuxième intégrale vaut μ, car au numérateur nous avons l'intégrale de Gauss qui vaut $\sqrt{2\pi}$

Nous aboutissons ensuite au résultat suivant : $E(X) = \mu$

4º) - La variance

La variance peut se calculer de la même manière, par intervention de la variable centrée normée U et en passant par l'intermédiaire de la fonction $\varphi(u)$. On aboutit à la relation

$$V(X) = \sigma^2$$

Ainsi les calculs précédents relient les paramètres de la loi de Laplace-Gauss, à son espérance mathématique, et sa variance.

Il s'en suit que si on utilise une variable centrée normée :

$$p_U(u) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-u^2}{2}\right)$$

(ci-joint la courbe représentative de cette fonction nº 1)

III - La loi de distribution de Cauchy

C'est une loi à deux paramètres, peu connue bien que assez utilisée.

1º) Formule de la loi de distribution :

$$p_X(x) = \frac{1}{\pi}\left(\frac{\sigma}{\sigma^2 + (x-\mu)^2}\right)$$

2º) Propriétés de cette loi

1 - Cette loi existe pour toute valeur de x.

2 - Elle est symétrique par rapport à la droite $x = \mu$. Par conséquent μ représente bien sa médiane.

Si on étudie la dérivée première de cette loi, on trouve une valeur proportionnelle à $\left[(\sigma^2 + (x-\mu)^2\right]^{-1}$

soit $\left[\sigma^2 + (x-\mu)^2\right]^{-2} 2(x-\mu)$

Le terme entre crochets ne peut pas s'annuler.

La dérivée a une valeur nulle lorsque $x = \mu$, ce qui montre que la médiane est également le mode de cette loi.

La dérivée seconde est proportionnelle à :

$$\left(\sigma^2 + (x-\mu)^2 - 4(x-\mu)^2\right)\left((x-\mu)^2 + \sigma^2\right)^{-3}$$

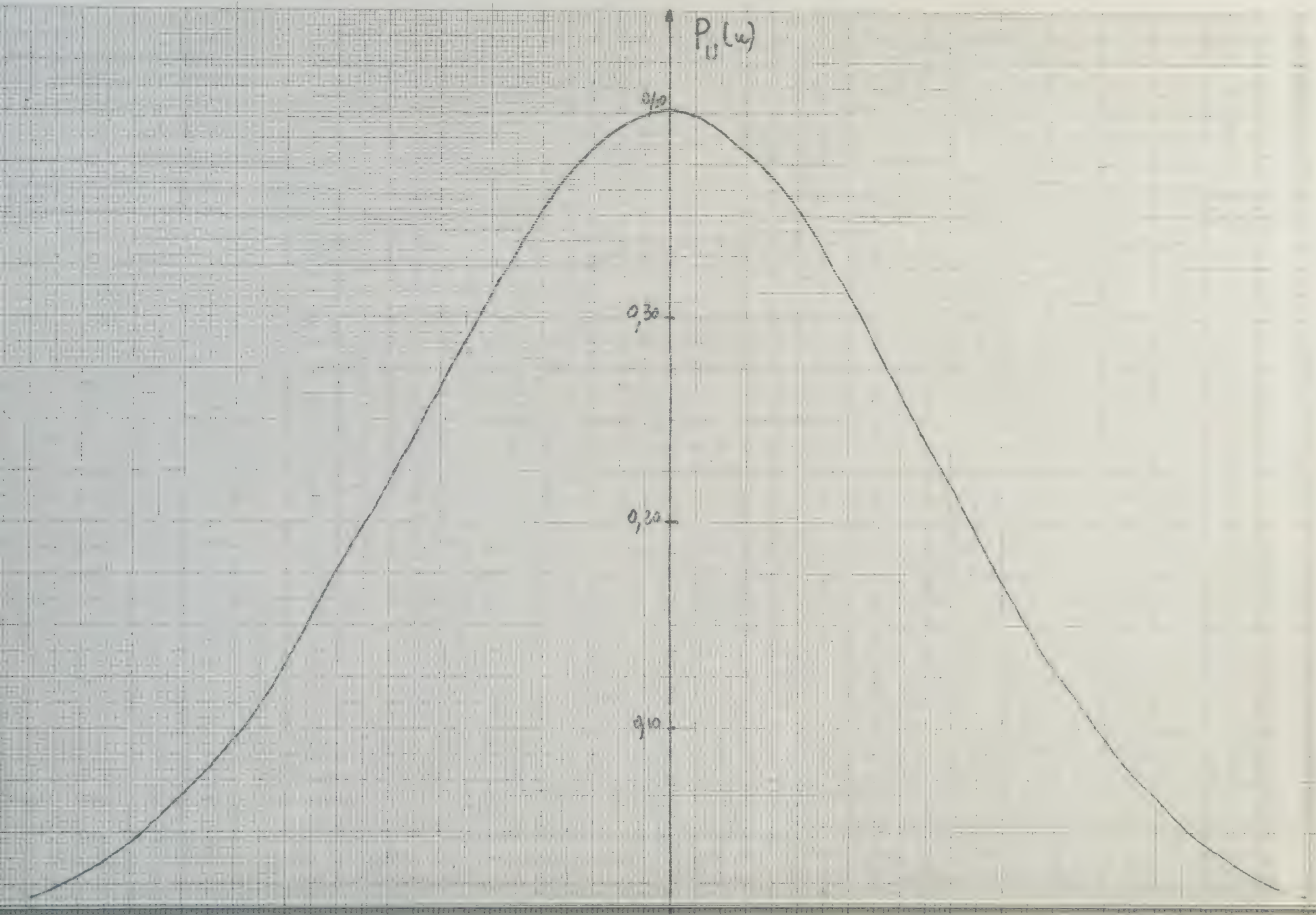
donc à $\sigma^2 - 3(x-\mu)^2$

valeur qui s'annule pour :

$$x = \frac{\mu - \sigma}{\sqrt{3}} \quad \text{et} \quad n = \frac{\mu + \sigma}{\sqrt{3}}$$

Ceci montre que cette courbe possède deux points d'inflexion séparés de la moyenne par la quantité $\frac{\sigma}{\sqrt{3}} = 0{,}577\sigma$. La forme de cette courbe est très analogue à celle de la loi de Gauss, mais les points d'inflexion de cette courbe sont plus rapprochés de la moyenne que dans la loi de Gauss.

Courbe représentative de la loi de Laplace - Gauss.



$$3 - \int_{-\infty}^{+\infty} p_x(x)\,dx = \frac{\sigma}{\pi} \int_{-\infty}^{+\infty} \frac{1}{\sigma^2 + (x-\mu)^2}\,dx$$

si on utilise la variable normée centrée $\frac{x-\mu}{\sigma}$, si l'on prend comme origine la médiane de X et σ comme unité de mesure :

$$x = u\,\sigma + \mu \qquad dx = \sigma\,du.$$

$$p_u\,du = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{1}{1+u^2}\,du = \frac{1}{\pi} \left[\text{arc tg } u\right]_{-\infty}^{+\infty}$$

4 - Espérance mathématique $E_{(X)}$ ou plus simplement $E_{(u)}$

$$E_{(u)} = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{u}{1+u^2}\,du = \frac{1}{2\pi} \left[\text{Log } (1+u^2)\right]_{-\infty}^{+\infty}$$

c'est une valeur <u>indéterminée</u>.

5 - Il est classique de montrer la même chose avec la variance.

La distribution de Cauchy possède donc une médiane, mais ni espérance mathématique, ni variance ou moment d'ordre plus élevé.

6 - Utilisation des propriétés de cette loi.

Soit une variable, "t", dite variable de Student. On peut la définir comme le quotient d'une variable aléatoire normale normée centrée par la racine carrée d'une variable distribuée suivant une loi de χ^2 à ν degrés de liberté (qui est la loi de la somme des carrés de ν variables aléatoires indépendantes mais toutes normales centrées et normées) divisé par son nombre de degrés de liberté, par la formule :

$$t = \frac{U}{\sqrt{\frac{\chi^2}{\nu}}} .$$

On parle alors de variable t à ν degrés de liberté.

La théorie montre que la loi de distribution de t est égale à :

$$\sqrt{\frac{1}{\pi}} \cdot \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma \frac{\nu}{2}} \cdot \frac{1}{\left(1 + \frac{u^2}{\nu}\right)\left(\frac{\nu+1}{2}\right)}$$

si $\nu = 1$ $\quad \Gamma\left(\frac{V+1}{2}\right) = \Gamma\,1 = 1$

$$\Gamma \frac{V}{2} = \Gamma \frac{1}{2} = \sqrt{\pi}.$$

La loi de distribution d'une variable de Student à 1 seul degré de liberté est donc $\frac{1}{\pi} \frac{1}{(1 + u^2)}$.

C'est une loi de Cauchy.

Cette définition permet de comprendre pourquoi, dans la suite de notre exposé, nous retrouverons la loi de Cauchy.

Nous avons à tenir compte des propriétés de la loi de Cauchy assez particulières, de n'avoir ni espérance mathématique, ni variance.

Ci-joint la courbe représentative de la loi de Cauchy (courbe nº 2).

Il est bien entendu que la courbe représentative de la loi de Cauchy monte moins haut que celle de la loi de Laplace-Gauss.

En effet, si u = 0 pour la loi de Gauss, on a la valeur $\frac{1}{\sqrt{2\pi}}$.

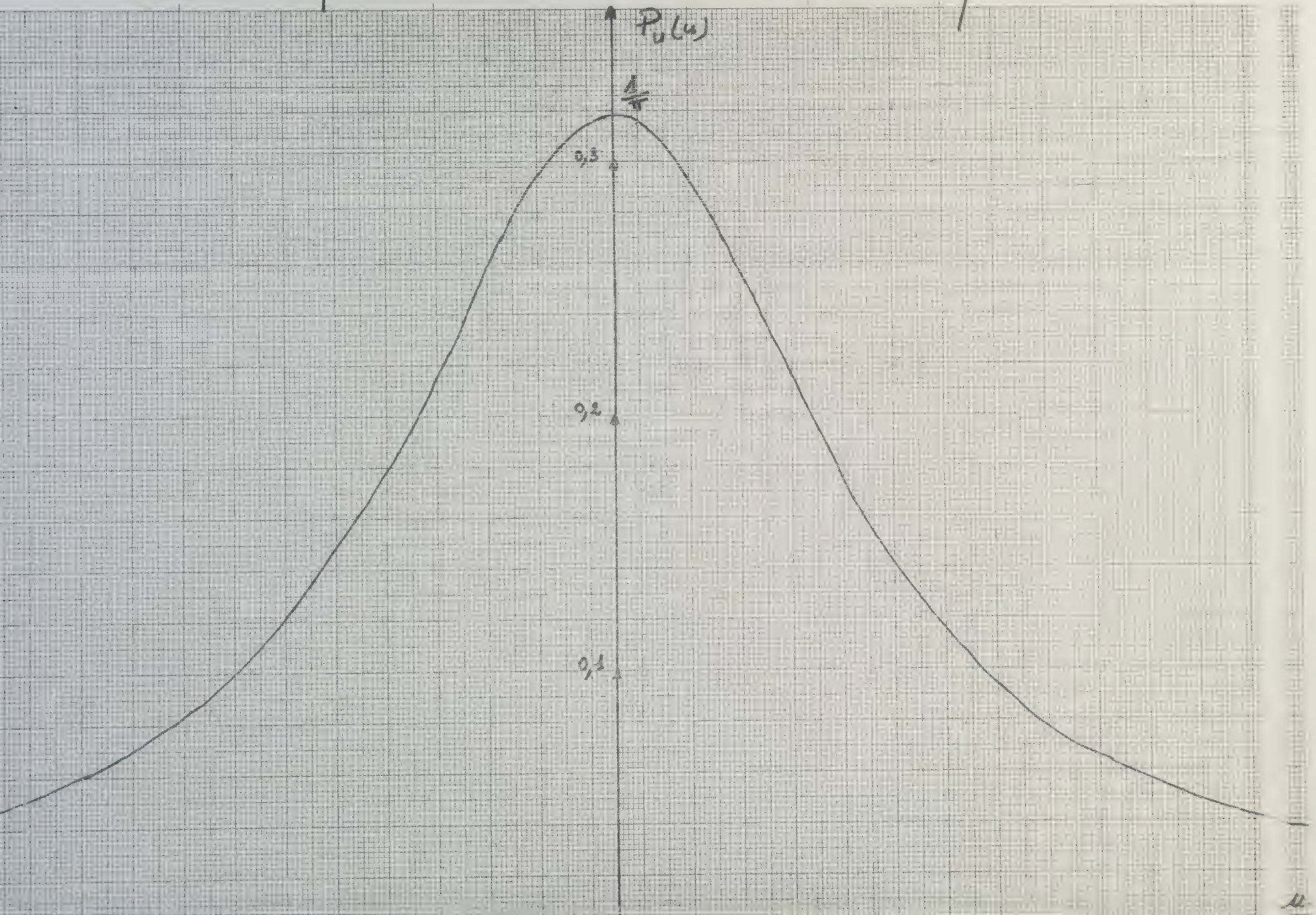
Pour la loi de Cauchy $\frac{1}{\pi}$.

Loi de Cauchy : tableau donnant p(u)

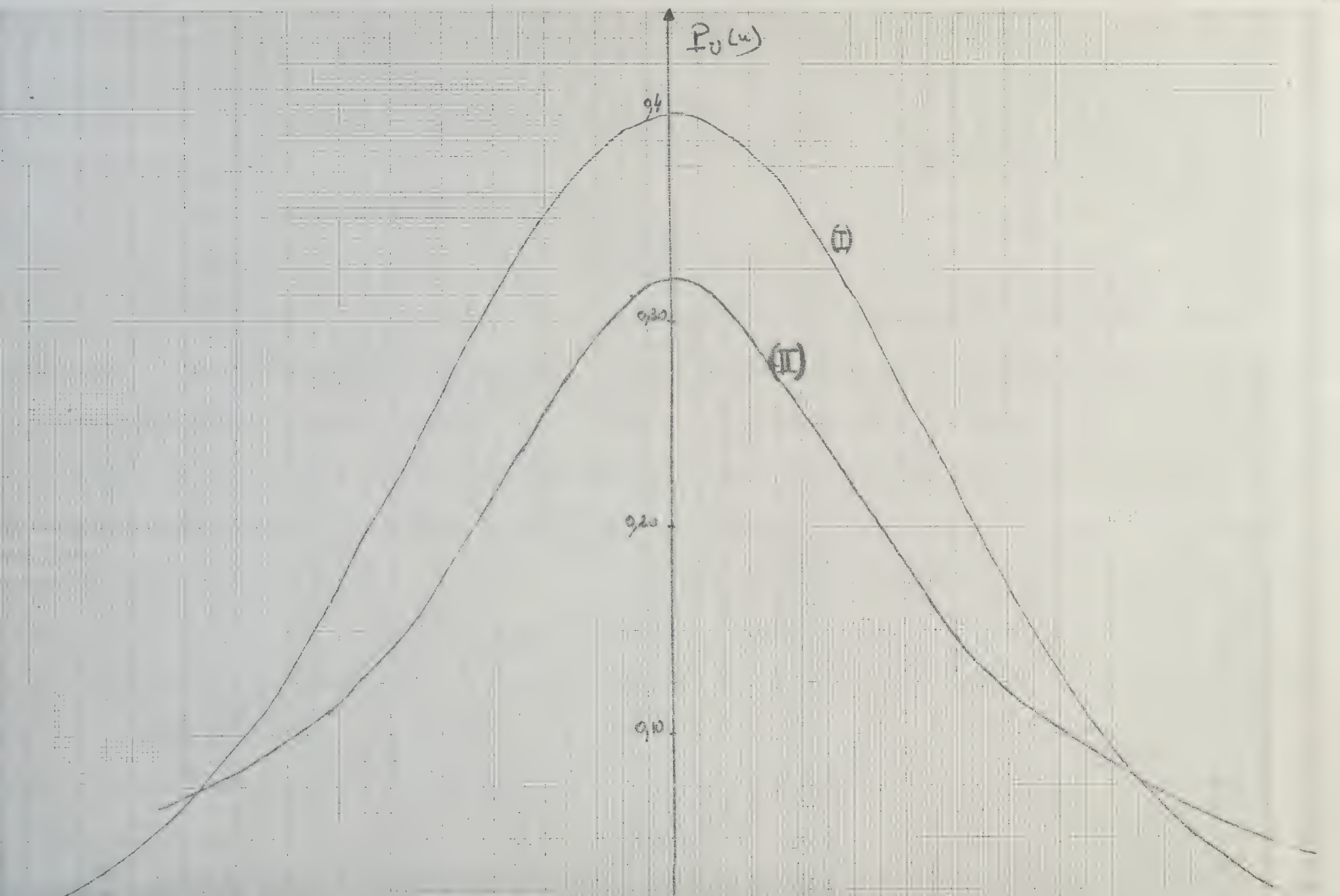
u	$\pm$ 0,25	$\pm$ 0,50	$\pm$ 0,75	$\pm$ 1
y	0,3	0,254	0,204	0,159
u	$\pm$ 1,25	1,50	$\pm$ 1,75	$\pm$ 2
y	0,124	0,098	0,078	0,064
u	$\pm$ 2,25	2,50	$\pm$ 2,75	$\pm$ 3
y	0,052	0,044	0,037	0,032

(voir courbe nº II)

Courbe Représentative de la loi de Cauchy (n:II)



Courbes représentatives de la loi de Laplace Gauss (1) et Ca...



CHAPITRE SECOND

FONCTIONS D'UNE SEULE VARIABLE ALEATOIRE - CAS PARTICULIER DE LA TRANSFORMATION LINEAIRE - CAS DE LA TRANSFORMATION NON LINEAIRE

1º) Fonctions d'une seule variable

Avant de nous lancer dans la description des fonctions de plusieurs variables, nous ferons une étude succinte du comportement des fonctions d'une seule variable. Nous verrons ainsi les difficultés auxquelles on se heurte, et les approximations que l'on est amené à faire dans ce cas simple.

Soit X, une variable aléatoire continue, susceptible de prendre x comme valeur numérique.

Nous avons défini, dans le chapitre précédent :

- l'espérance mathématique de X: $E(X) = \int_a^b x\, p_X(x)\, dx$ (bornes de l'intervalle de définition a et b).
- la variance de X: $V(X) = \int_a^b (x - E(X))^2\, p_X(x)\, dx$.

Condidérons maintenant une variable Y qui est définie numériquement par une valeur y, liée à la valeur numérique de x par la loi :

(I) $$\boxed{y = \varphi(x)}$$

Nous ne considérerons que des fonctions φ monotones croissantes et biuniveques

Par conséquent la dérivée $\varphi'(x)$ est positive.

A cause de la bijection, on peut faire correspondre à une valeur donnée y, une valeur x

(II) $$\boxed{x = \psi(y)}$$

qui a les mêmes qualités, ainsi que $\psi'(y)$.

La monotonie de la fonction n'est pas indispensable mais lorsqu'elle n'est pas réalisée, les calculs deviennent rapidement plus complexes, comme le montre l'expérience de l'étude de la distribution du χ^2.

I - Formules fondamentales

prob $(X \leq x)$ = prob $(Y \leq y)$ pour tout couple de x et y unis par (I) et (II).

Par suite, en différenciant et en appelant $p_Y(y)dy$ la probabilité pour que $y < Y < y + dy$, on peut écrire :

$$\text{(III)} \qquad \boxed{p_X(x)dx = p_Y(y)dy.}$$

La probabilité pour que X soit comprise entre x et x + dx, est identique à celle pour que Y soit comprise entre y et y + dy.

Appliquons les règles de changement de variable et remplaçons dans le membre de gauche x par $\psi(y)$ et x'par $\psi'y\,dy$.

$$p_Y(y) = \left(p_X(x)\right)_{x = \psi(y)} \cdot \psi'(y) .$$

Il n'y a aucune raison pour que p_X et p_Y aient la même forme mathématique.

Ceci veut dire, par exemple : il n'y aucune raison pour qu'une loi de Laplace-Gauss relative à X soit transformée en une loi de Laplace-Gauss, relative à Y.

II - Espérance mathématique de Y. $E_{(Y)}$

$$E_{(Y)} = \int_c^d y\, p_Y(y)\, dy.$$

où les limites d'intégration c et d se déduisent de a et b par les règles usuelles des intégrales définies.

Mais on peut tenir compte de la transformation

$$E_{(Y)} = \int_a^b \varphi(x)\, p_X(x)\, dx, \quad \text{d'après la relation (III.)}$$

Ce n'est que dans des cas particulièrement simples, que la relation qui unit $E_{(X)}$ et $E_{(Y)}$ est simple elle-même, comme nous aurons l'occasion de le voir.

III - Variance de Y : $V_{(Y)}$.

$$V_{(Y)} = \int_c^d (y - E_{(Y)})^2 \, p_Y(y) \, dy.$$

Développons pour simplifier les calculs :

$$V_{(Y)} = \int_c^d y^2 \, p_Y(y) \, dy - 2 \int_c^d y \, E_{(Y)} \, p_Y(y) dy + \int_c^d (E_{(Y)})^2 \, dy$$

Mais le dernier terme peut s'écrire :

$$\int_c^d (E_{(Y)})^2 dy = (E_{(Y)})^2 \int_c^d dy = (E_{(Y)})^2$$

De même, pour le deuxième terme :

$$2 \, E_{(Y)} \int_c^d y \, p_Y(y) \, dy = 2 (E_{(Y)})^2$$

$$V_{(Y)} = \int_c^d y^2 \, p_Y(y) dy - (E_{(Y)})^2$$

$$V_{(Y)} = \int_a^b \varphi^2(x) \, p_X(x) dx - (E_{(Y)})^2$$

Là encore, les relations qui relient $V_{(X)}$ et $V_{(Y)}$ sont simples seulement dans de rares cas.

2º) Cas particulier de la transformation linéaire.

C'est le cas particulièrement simple où

$y = \alpha + \beta x$ (α et β sont des quantités certaines).

L'application des formules précédentes nous donne immédiatement :

1 - Espérance mathématique :

$$E_{(Y)} = \int_a^b (\alpha + \beta x) \, p_X(x) \, dx.$$

$$E_{(Y)} = \int_a^b \alpha \, p_X(x) \, dx + \int_a^b \beta x \, p_X(x) \, dx$$

Par définition, la première intégrale vaut α, la deuxième intégrale vaut $\beta E_{(X)}$.

Nous trouvons la formule remarquable :

(IV) $\boxed{E_{(Y)} = \alpha + \beta E_{(X)}}$

2 - Variance

Quant à la variance, elle se calcule fort simplement :

$$V_{(Y)} = \int_a^b (\alpha + \beta x)^2 \, p_X(x) \, dx = (E_{(Y)})^2$$

en développant le carré nous trouvons la somme de trois intégrales.

La première vaut $\int_a^b \alpha^2 \, p_X(x) \, dx = \alpha^2$

La seconde $\int_a^b 2 \alpha \beta x \, p_X(x) \, dx = 2 \alpha\beta \, E(X)$

La troisième $\int_a^b \beta^2 x^2 \, p_X(x) \, dx$ comme d'après la formule de KÖNIG

$$V_{(X)} = \int_a^b x^2 \, p_X(x) \, dx - (E_{(X)})^2$$

on peut écrire que cette troisième intégrale vaut :

$$\beta^2 \, (V_{(X)} + (E_{(X)})^2)$$

Par suite l'intégrale complète vaut :

$$\alpha^2 + 2 \alpha\beta \, E(X) + \beta^2 (V_{(X)} + E_{(X)})^2$$

En remplaçant E_y par sa valeur $\alpha + \beta \, E_{(X)}$ on trouve

$$V_{(Y)} = \alpha^2 + 2 \alpha \beta E_{(X)} + \beta^2 (V_{(X)} + (E_{(X)})^2)$$

$$- \alpha^2 - 2 \alpha \beta \, E_{(X)} - \beta^2 (E_{(X)})^2$$

(V) Donc $\boxed{V_{(Y)} = \beta^2 V_{(X)}}$ résultat fort simple.

3 - Loi de Gauss et Cauchy

- Il est évident, d'après les formules de la transformation linéaire, qu'une variable X, suivant la loi de Laplace-Gauss , va être transformée en une variable Y, suivant aussi la loi de Laplace-Gauss.

- Il en est de même pour la loi de Cauchy.

En effet, :

$$y = \alpha + \beta x \qquad x = \frac{y - \alpha}{\beta}$$

C'est pourquoi on considère beaucoup en Biologie, un facteur qui est insensible au changement d'unités: le coefficient de variation:

$$\frac{\sqrt{V(X)}}{E(X)}$$

3º) Fonctions dont la transformation est non linéaire.

Les résultats donnés ici, ne pourront être utilisés que dans des cas très particuliers.

Dans la pratique on rencontre le cas où $y = \varphi(x)$ est développable en série de Taylor autour de la valeur $E(X)$.

Cette série de Taylor peut se limiter au terme du premier ordre en x.

Elle s'écrit :

$$y = \varphi E(X) + (x - E(X))\,\varphi' E(X)$$

$$y = \varphi E(X) - E(X)\,\varphi' E(X) + x\,\varphi' E(X)$$

Appliquons les formules précédentes (IV) et (V). Nous pourrons écrire :

$$E_Y = \varphi E(X) + 0$$

Autrement dit l'espérance mathématique de Y est égale à la fonction de l'espérance mathématique de X

si $y = \frac{1}{x}$ on a donc $E(Y) = \frac{1}{E(X)}$

et ainsi de suite.

D'autre part, pour la variance :

$$V(Y) = \left[\varphi' E(X)\right]^2 V(X)$$

La formule la plus intéressante est celle concernant l'espérance mathématique de Y.

Il faut avoir présente à l'esprit, la restriction du développement de Taylor, ce qui signifie que X varie peu autour de son espérance mathématique.

Dans la réalité, le domaine de variation de X est restreint alors que le domaine théorique peut aller de $-\infty$ à $+\infty$.

Ainsi la valeur de $E(\frac{1}{X})$ est différente de celle de $\frac{1}{E(X)}$, ce que nous verrons dans le chapitre suivant.

L'égalité n'est vraissemblable que dans un domaine restreint.

D'autre part, si la formule de Taylor est applicable, la distribution de Laplace-Gauss et Cauchy sont conservées puisque la formule de Taylor limitée à ses premiers termes n'est pas autre chose que linéaire. Donc, lorsque les valeurs de X varient peu autour d'une valeur moyenne et si X est distribuée suivant une loi de Laplace-Gauss, une fonction arbitraire de X est aussi distribuée suivant la loi de Laplace-Gauss. Ceci explique que pour Heller, les variables comme la taille, la surface corporelle, et le poids d'un sujet, sont toutes les trois distribuées normalement.

CHAPITRE TROISIEME

Fonctions de plusieurs variables

Nous considérerons une variable aléatoire Z dont la valeur numérique est une fonction des valeurs numériques d'autres variables aléatoires.

En principe, on peut considérer un nombre quelconque de variables aléatoires, et cela pourra nous arriver dans certains cas.

En fait, le cas le plus classique est celui où

$$Z = \varphi(X, Y)$$

Ainsi il n'y a que deux variables aléatoires X et Y d'espérance mathématique $E_{(X)}$ et $E_{(Y)}$, de variance V(X) et V(Y), et de fonction de distribution $P_X(x)$ et $P_Y(y)$.

1º) Relations qui unissent deux variables aléatoires.

Les deux variables aléatoires peuvent être stochastiquement indépendantes ou non.

Souvent, ces deux variables aléatoires ne sont pas stochastiquement indépendantes.

Cela se traduit par le fait que la probabilité pour que $y < Y < y + dy$ est affectée par la valeur x.

On parle alors de probabilité conditionnelle de Y étant donné la valeur de X.

Ceci s'écrit prob (Y/ X)

Il existe naturellement une prob (X/Y).

- Nous considérons le cas où les deux variables aléatoires X et Y sont stochastiquement indépendantes.

Ceci se traduit par l'équation :

prob (Y/X) = $p_Y(y)$.

La valeur de X n'apparaît pas comme paramètre dans ces équations, autrement dit tout se passe comme si Y était considérée comme seule.

Bien entendu on peut écrire aussi :

$$p\ (X/Y) = p_X(x).$$

Notons aussi la formule donnant la probabilité pour que X et Y aient elles-mêmes les valeurs numériques x et y, dans le cas de la dépendance.

$p_X(x)$ prob (Y/X) dx dy

Cette formule peut être remplacée par l'expression :

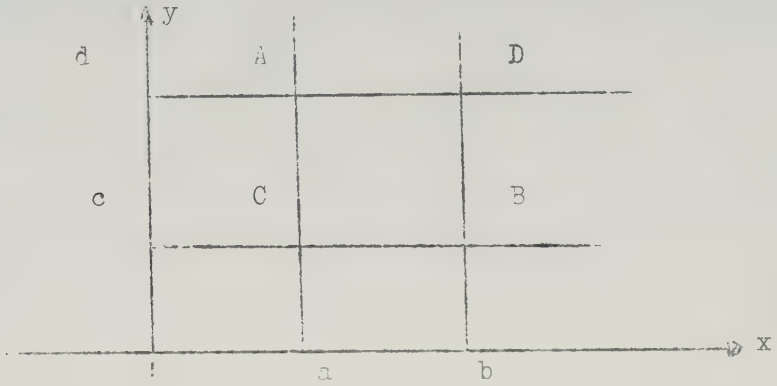
$p_Y(y)$ prob (X/Y) dx.dy.

2º) Covariance de X et Y et coefficient de corrélation

Soient a et b les limites de variation de X
c et d les limites de variation de Y

a) On appelle covariance de X et Y : "C (X, Y)" la quantité $\iint_D (x - E(X))$ $(y - E(Y))\ p_X(x)$ prob (X/Y) dx dy.

Le domaine D est un domaine rectangulaire limité par les points d'abcisse a et b.



b) On définit aussi le coefficient de correlation entre X et Y noté $\rho(X,Y)$

$$\rho(X,Y) = \frac{C(X, Y)}{\sigma_X \cdot \sigma_Y}$$

On démontre que ρ est inférieur à 1.

3º) Covariance et indépendance stochastique.

Dans le cas de l'indépendance stochastique la formule qui donne la covariance est :

$$C(X , Y) = \iint_D (x - E(X))\ (y - E(Y))\ p_X(x)\ \ p_Y(y)\ dx.dy.$$

D'après les propriétés des intégrales doubles étendues à des domaines rectangulaires.

$$C(X , Y) = \int_a^b (x - E(X))\ p_X(x)\ dx \int_c^d (y - E(Y))\ p_Y(y)\ dy.$$

Or, d'après la définition de l'espérance mathématique chacune de ces intégrales est nulle.

Donc la covariance ou le coefficient de correlation de deux variables aléatoires stochastiquement indépendantes est nul.

Ce théorème n'admet pas de réciproque.

Ainsi on ne peut rien dire sur la dépendance ou l'indépendance de deux variables aléatoires dont la covariance est nulle.

Nous verrons que dans certains cas, deux variables aléatoires, liées, ont une covariance nulle.

Dans la formule donnant la covariance :

les termes $P(X)$, $P(Y)$ sont positifs

$x - E(X)$

$y - E(Y)$ sont positifs ou négatifs.

Ainsi la covariance est positive si X ou Y ont une valeur élevée en même temps.

Elle est négative dans le cas contraire.

Le fait que la réciproque ne soit pas vraie a beaucoup gêné les chercheurs. Ainsi ils ont cherché un critère permettant d'affirmer l'indépendance stochastique de X et Y au seul vu des résultats.

FELLER fait une discussion du problème, page 133 du tome II - Edition 1966.

Théorème : La condition nécessaire et suffisante pour que deux variables aléatoires X et Y centrées soit stochastiquement indépendantes, est que pour toutes fonctions u, v nulles à l'extérieur de l'intervalle fini,

$$E\left(u(X), v.(Y)\right) = E\left(u(X)\right).\ E\left(v(Y)\right)$$

4º) Covariance et espérance mathématique du couple (X, Y)

Par définition la covariance n'est pas autre chose que l'espérance mathématique de la variable aléatoire définie par les valeurs numériques

$$\left(x - E(X)\right)\left(y - E(Y)\right)$$

Ceci entraine des conséquences nombreuses, en développant

$$x\,y - E(X)\,y - E(Y)\,x + E(X)\,E(Y) = \left(x - E(X)\right)\left(y - E(Y)\right)$$

D'où $\boxed{E(X, Y) - E(X)\ E(Y) = C(X,Y)}$

Ceci permet de démonter que

$$E\left(\frac{1}{X}\right) > \frac{1}{E(X)}$$

inégalité étendue à tout domaine de variation.

Considérons la variable $Y = \frac{1}{X}$

C (X, Y) est < 0 si X ↗ Y ↘

Le domaine de variation sera pris avec :

- une borne supérieure infinie
- une borne inférieure nulle.

X.Y = 1 quantité certaine dont l'espérance mathématique vaut 1. Donc

$$1 - E(X).E(Y) \leq 0$$

$$E(X).E(Y) \geq 1 .$$

$E(Y) > \frac{1}{E(X)}$ l'égalité ne sera réalisée que dans le cas où l'approximation pour la série de Taylor est valable.

Cette formule permet en outre de relier, la covariance des variables aléatoires X' et Y' (liées à X et Y elles-mêmes).

Posons :

$$X' = \alpha + \beta X$$

$$Y' = \gamma + \delta Y.$$

$$E(X') = \alpha + \beta E(X)$$

$$E(Y') = \gamma + \delta E(Y).$$

Calculons le produit $E(X') E(Y')$: $\alpha \gamma + \beta \gamma E(X) + \alpha \delta E(Y) + \beta \delta E(X) E(Y)$

Or $X'Y' = \alpha\gamma + \beta\gamma x + \alpha\delta y + \beta\delta xy$

Donc $E(X'Y') = \alpha \gamma + \beta \gamma E(X) + \alpha \delta E(Y) + \beta \delta E(X.Y)$

$$C(X', Y') = \beta \delta \left(E(X,Y) - E X. E Y \right)$$

$$C(X',Y') = \beta \delta C(X,Y)$$

Ainsi, si nous faisons un changement d'unité consistant à mesurer la grandeur x, avec une unité h fois plus grande, et y, avec une unité k fois plus grande, la covariance des grandeurs x, y vaut l'ancienne covariance multipliée par la quantité h . k

CHAPITRE QUATRIEME

Somme de variables aléatoires

Il s'agit évidemment de sommes algébriques. Nous considérerons la variable aléatoire Z définie à partir des variables aléatoires X et Y par la relation numérique.

$$Z = X + \varepsilon Y \qquad \varepsilon = \pm 1 .$$

Nous calculerons l'Espérance mathématique de Z:E(Z), sa variance:V(Z).

Nous envisagerons rapidement le cas de la transformation éventuelle des lois de distribution.

Enfin, nous élargirons au cas d'un nombre de variables aléatoires supérieur à deux.

I - Généralités

Nous remplacerons la formule faisant intervenir les probabilités conditionnelles, par la relation unique p(x,y) qui donne la probabilité pour que le couple de variables aléatoires X, Y aient en même temps les valeurs numériques x et y.

Autrement dit, la probabilité pour que l'on ait en même temps

$$\begin{matrix} x < X < x + dx \\ y < Y < y + dy \end{matrix} \quad \text{s'écrira } p(x,y)\, dx\, dy.$$

Dans ces conditions, calculons E(Z) espérance mathématique de Z

$$E(Z) = \iint_D (x + \varepsilon y)\ p\ (x, y)\, dx\, dy$$

$$E(Z) = \underbrace{\iint_D x.p\ (x, y)\, dx\, dy}_{(1)} + \varepsilon \underbrace{\iint_D y\ p\ (x, y)\, dx\, dy}_{(2)}$$

La première intégrale n'est pas autre chose que l'intégrale $\int_a^b x\ p_X\ (x)\, dx$.

Elle exprime en effet la quantité intégrée étendue dans un domaine comportant toutes les valeurs de y et toutes les valeurs de x c'est-à-dire $\int_a^b x\, dx \int_c^d p\ (x, y)\, dy$

Or $\int_c^d p\ (x,y)\, dy$ vaut $\mu_X(x)$

Par suite, l'intégrale double (1) exprime E(X), la deuxième E(Y).

$$\text{d'où} \qquad E(Z) = E(X) + \varepsilon\, E(Y).$$

Remarque :

- Le raisonnement reste valable pour un nombre de variables aléatoires plus grand que deux, liées ou non.
- Les espérances mathématiques s'ajoutent ou se retranchent suivant le signe de ϵ.

V(Z) variance de Z

$$(Z - E(Z))^2 = ((X + \epsilon Y) - (E(X) + \epsilon E(Y)))^2$$

$$= (X - E(X) + \epsilon (Y - E(Y)))^2 =$$

$$(X - E(X))^2 + 2\epsilon(X - E(X))(Y - E(Y)) + \epsilon^2 (Y - E(Y))^2$$

en prenant les espérances mathématiques :

$$E(X - E_{(X)})^2 + 2 \epsilon E (X - Ex)(Y - E_{(Y)}) + \epsilon^2 E(Y - E_{(Y)})^2$$

D'où :

$V(Z) = V(X) + 2C(X,Y) + V(Y)$ car $\epsilon^2 = 1$.

On voit apparaître dans cette formule, la covariance de X et Y.

Remarque

1º) Si X et Y sont stochastiquement indépendantes la covariance est nulle, comme nous l'avons vu.

Donc $V(Z) = V(X) + V(Y)$, les variances s'ajoutent toujours.

Si la covariance a une valeur différente de 0, nous avons $V(Z) \neq V(X) + V(Y)$.

2º) La formule peut être généralisée à plus de deux variables aléatoires.

3º) Si $Z = \alpha X + \beta Y$ et si X et Y sont stochastiquement indépendantes, un calcul analogue à celui du chapitre précédent montre que

$$V(Z) = \alpha^2 V(X) + \beta^2 V(Y) + \alpha \beta C(X,Y)$$

car $C(\alpha X + \beta Y) = \alpha \beta C(X, Y)$.

On montre aussi que

$$E(Z) = \alpha E(X) + \beta E(Y)$$

On peut généraliser si φ est développable en série de Taylor, limitée au premier ordre par une formule du type :

$$\varphi(x, y) = \varphi(E(X), E(Y)) + (x - E(X)) \varphi'_x (E(X), E(Y)) + (y - E(Y)) \varphi'_y (E(X), E(Y))$$

ceci conduit à la formule remarquable

$$\boxed{E(Z) = \varphi(E(X), E(Y))}$$

II - Exemples d'application.

Exemple I

Supposons que X et Y soient deux variables centrées, cela signifie que E(X) et E(Y) sont nulles.

$$E(X) = E(Y) = 0 \qquad \text{et} \qquad E(X)^2 = E(Y)^2$$

Considérons deux variables Z_1 et Z_2 dont les valeurs numériques sont définies par :

$$z_1 = x + y.$$

$$E(Z_1) = E(X) + E(Y) = 0$$

$$z_2 = x - y$$

$$E(Z_2) = E(X) - E(Y) = 0.$$

Les variables Z_1 et Z_2 sont liées.

Cherchons la valeur de la covariance C

$$C(Z_1, Z_2) = E(Z_1 Z_2) - E(Z_1) E(Z_2)$$

Donc, puisque $E(Z_1) = E(Z_2) = 0$

$$C(Z_1, Z_2) = E(Z_1 Z_2) = E((X + Y)(X - Y))$$

$$Z_1 Z_2 = (x + y)(x - y)$$

$$E(Z_1 Z_2) = E(X^2) - E(Y^2)$$

Ces quantités sont égales, leur différence vaut 0.

Par suite, la covariance de la somme et de la différence de deux variables aléatoires identiques est nulle quelle que soit la loi de distribution de ces deux variables.

On voit sur cet exemple que deux variables aléatoires liées peuvent avoir une covariance nulle.

Exemple II

Il est très important, dans la pratique, parce qu'il permet d'étudier le biais des estimateurs de la variance. Partons de la définition de la variance.

$$V(X) = E(X - \mu)^2$$

$$V(X) = E(X^2) - \mu^2.$$

Ceci peut s'écrire :

$$E(X^2) = V(X) + \mu^2$$

Cette forme montre qu'évidemment $E(X^2)$ est différent de la valeur μ^2 que donnerait l'application du développement de Taylor.

III - Différentes formes de distribution

Soient X et Y deux variables stochastiquement indépendantes, ainsi que Z qui est la somme de X et Y.

La loi de distribution de X, Y, Z sera simple. Dans ce cas la formule qui donne la probabilité pour que Z ait une valeur comprise entre z et z + dz est celle-ci : $P_Z(Z)\ dz$.

Cette probabilité est égale au produit de la probabilité pour que X ait une valeur comprise entre deux valeurs x et x + dx (où x a une valeur quelconque imposée) par la probabilité pour que Y soit comprise entre Z - x et Z + dz - x.

La première de des probabilités vaut $\int_x^{x + dx} p_X(x)\ dx$

La seconde vaut $\int_{z - x}^{z + dz - x} p_Y(y)\ dy.$

D'après le théorème de la moyenne, cette intégrale vaut :

$$dz.p_Y\ (z - x)$$

(Mais, ce faisant, nous avons imposé à x une valeur, alors qu'il peut avoir une valeur quelconque dans l'intervalle ab de variation de x).

Nous aboutissons donc à la formule

$$p_Z(z)\ dz = dz \int_a^b p_X(x)\ \ p_Y(z - x)\ dx.$$

Ceci nous donne :

$$\boxed{p_Z(z) = \int_a^b p_X(x)\ p_Y(z - x)\ dx.}$$

Cette formule est fondamentale et se retrouve dans la théorie de la fiabilité.

1º) Cas de la loi de Laplace Gauss.

On supposera X et Y centrées, de variances différentes σ_X^2 et σ_Y^2.

$$p_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma_X} \exp\left(-\frac{x^2}{2\sigma_X^2}\right).$$

$$p_Y(y) = \frac{1}{\sqrt{2\pi}\,\sigma_Y} \exp\left(-\frac{y^2}{2\sigma_Y^2}\right).$$

$$p_Y(z - x) = \frac{1}{\sqrt{2\pi}\,\sigma_Y} \exp\left(-\frac{(z-x)^2}{2\sigma_y^2}\right)$$

Nous allons développer le carré de l'argument de l'exponentielle qui s'écrira :

$$\exp\left(\frac{-z^2}{2\sigma_y^2} + \frac{z\,x}{\sigma_y^2} - \frac{x^2}{\sigma_y^2}\right)$$

Donc :

$$p_Z(z) = \frac{1}{2\pi\,\sigma_x\,\sigma_y} \int_{-\infty}^{+\infty} \exp\left(\frac{-x^2}{2\sigma_x^2}\right). \exp\left(\frac{-(z-x)^2}{2\sigma_y^2}\right) dx$$

Le produit des deux exponentielles se met sous la forme d'une exponentielle dont l'argument vaut :

$$-\frac{x^2}{2\sigma_x^2} - \frac{x^2}{2\sigma_y^2} + \frac{z\,x}{\sigma_y^2} - \frac{z^2}{2\sigma_y^2}$$

D'après ce que nous savons sur la variance, on peut écrire, si nous appelons σ_z^2 la variance de z et σ_x^2 et σ_y^2 celle de X et Y.

$$\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2$$

En appliquant les règles du changement de variable, nous trouverons, si nous effectuons le changement :

$$u = \frac{\sigma_Z}{\sigma_x.\sigma_y}\left(x - z\,\frac{\sigma_x^2}{\sigma_y^2}\right)$$

$$p_Z(z) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(\frac{-z^2}{2\sigma_z^2}\right) \frac{\sigma_x \cdot \sigma_y}{\sigma_z} \int_{-\infty}^{+\infty} \exp\left(-\frac{u^2}{2}\right) du$$

$$p_Z(z) = \frac{1}{\sqrt{2\pi}\,\sigma_z} \exp\left(\frac{-z^2}{2\sigma_z^2}\right)$$

Nous avons donc encore, avec Z une loi de Laplace-Gauss centrée.

Si X et Y ne sont pas centrées, le calcul subsiste avec un changement de variable préalable destiné au centrage et donne un bon résultat.

Théorème

La somme de deux variables normales stochastiquement indépendantes est une variable normale.

Généralisation à un nombre quelconque de variables : cette généralisation est possible.

Réciproque ou théorème de Lévy-Cramer

Si une variable aléatoire est distribuée normalement et si l'on sait quelle est la somme de 2 variables aléatoires stochastiquement indépendantes, celles-ci sont normalement distribuées.

La démonstration en est difficile.

Nous rapprocherons ces théorèmes du fameux théorème fondamental de la statistique : le théorème central limite, démontré en 1901 par LIAPOUNOV.

La distribution de plusieurs variables en nombre n stochastiquement indépendantes tend vers la loi normale lorsque n tend vers l'infini.

Sur les conditions générales, nous en retiendrons deux:

1) La variance de la somme ne tend pas vers l'infini.

2) La variance de chaque variable composante par rapport à la variance de la somme tend vers 0.

Nous remarquerons la différence entre le théorème relatif à des composantes normales, où il y a une restriction sur leur nombre, et le théorème central limite où x est grand.

2º) Loi de Gauchy

Soient deux variables de Cauchy définies par les lois :

$$p_X(x) = \frac{1}{\pi(1 + x^2)} \qquad p_Y(y) = \frac{1}{\pi(1 + y^2)}$$

et une troisième définie par

$$z = x + y.$$

Un calcul facile mais long montre que :

$$p_Z(z) = \frac{1}{\pi\,(1 + z^2)}$$

La somme de deux variables de Cauchy, indépendantes et de même paramètre, est une variable de Cauchy de même paramètre.

Si on fait la somme de n variables de Cauchy de même paramètre, cette somme est distribuée suivant une loi de Cauchy de même paramètre.

Remarque

La loi de Cauchy semble mettre le théorème central limite en défaut.

Mais nous savons que le théorème central limite nécessite des hypothèses sur la variance des composantes.La loi de Cauchy ne possédant pas de variance, ce théorème ne lui est pas applicable. Le paradoxe n'est donc qu'apparant.

Le théorème central limite ne nécessite pas que les distributions des variables composantes soient identiques. Mais si tel est le cas, il se simplifie et son résultat est valable pourvu que moyenne et variance de ces distributions existent. Sous cette forme on voit qu'il ne s'applique pas à la loi de Cauchy.

Pour être strict, la démonstration que la moyenne des échantillons d'une population suit la loi de Laplace-Gauss n'apporte aucun renseignement sur la normalité éventuelle de cette population.

On sait seulement que la loi des sujets qui composent un échantillon est la même pour tous.

Autres exemples.

Les distributions symétriques

Dire que la distribution d'une variable aléatoire est symétrique veut dire que :

Prob. $x < X < x + dx$ = Prob $-x-dx < X < -x$ ou dx est une quantité déterminée.

On a donc :

$$P_X(x) = P_X(-x)$$

$$E(X) = 0 \quad \text{naturellement.}$$

Considérons les variables aléatoires X et Y, dont l'une par exemple Y est symétrique, et considérons aussi les deux variables Z_1 et Z_2 dont les valeurs numériques sont $z_1 = x + y$

$$z_2 = x - y.$$

D'après les formules générales

$$P_{Z_1}(z_1) = \int_A^B p_X(x)\, p_Y(z_1 - x)\, dx$$

$$p\, Z_2(z_2) = \int_c^d p_X(x)\; p_Y(-(z_2 - x))\; dx.$$

A cause de la symétrie des Y, les 2 éléments différentiels sont égaux car

$$p_Y\,(z_2 - x) = p_Y\,-(z_1 - x)$$

On a le résultat suivant :

lorsque une des deux variables aléatoires composantes est symétrique, les distributions de Z_1 et Z_2 sont les mêmes.

CHAPITRE CINQUIEME

Produit de deux variables stochastiquement indépendantes.

Les résultats que nous pourrons obtenir seront encore plus fragmentaires que dans le chapitre précédent.

Espérance mathématique.

La formule fondamentale est la suivante :

$$C(X, Y) = E(X.Y) - E(X)\,E(Y)$$

$$E(X.Y) = C(X, Y) + E(X).\,E(Y).$$

Nous considérerons la variable Z dont la valeur numérique est $z = xy$.

$$E(Z) = E(X.Y)$$

si X et Y sont stochastiquement indépendantes

$$\boxed{E(Z) = E(X).\,E(Y).}$$

Variance

On ne peut rien dire de simple sur cette valeur.

Distribution du produit (X.Y)

Rien de simple n'a été écrit à ce sujet dans le peu de littérature existant.

SORIN et THIONET rattachent cette distribution aux fonctions de Bessel qui sont racines de l'équation différentielle du deuxième ordre, où z est une variable complexe.

$$z^2\,y'' + z\,y' - z^2 y = 0$$

équation différentielle qui admet deux solutions.

$J_0(z)$ et $K_0(z)$.

La seule qui intervient est $K_0(z)$

$$K_0(X) = \int_0^{\infty} \frac{\cos t(x)\ dt}{\sqrt{t^2 + 1}} \qquad (x \text{ réel} > 0)$$

Le calcul de SORIN et THIONET repose sur l'hypothèse que les variables X et Y sont stochastiquement indépendantes et il est mené de pair avec le calcul de la loi de distribution X + Y et X - Y.

Les variables X et Y sont supposées normales, et d'après SORIN et THIONET stochastiquement indépendantes. Or nous savons qu'une covariance nulle n'entraine pas automatiquement l'indépendance stochastique. En fait JIBBONS : "Mutually exclusive events, indépendance and zéro correlation" American Statistician - 1968 - 22 - nº 5 - p.31 et 32, donne un exemple relatif à une distribution discontinue, il est vrai, où la somme et la différence ne sont pas stochastiquement indépendantes. C'est pour cette raison que nous ne détaillerons pas le calcul de SORIN et THIONET, assez laborieux, fondé sur l'utilisation des équations caractéristiques et transformées de Laplace et nous donnerons le résultat :

$$p_Z(z) = \frac{1}{2\pi} K_0(|z|).$$

CHAPITRE SIXIEME

RAPPORT DE DEUX VARIABLES

En ce qui concerne la variable Z définie numériquement à partir de X et Y par une loi du type :

$$z = \frac{x}{y}$$

les résultats, non triviaux relatifs à l'espérance mathématique E(Z) et qui conduisent à :

$$E(Z) = \frac{E(X)}{E(Y)}$$

lorsqu'on se tient près de "espérance mathématique de X et Y" permettant le développement en série de TAYLOR, remontent au début du siècle et constituent des outils importants de la statistique appliquée par certains de leurs aspects. Ils concernent essentiellement :

- La variable "F" qui est utilisée dans le test de Snédécor relatif à l'homoscédasticité de deux échantillons issus d'une même population normale.
- La variable "t" de Student, qui sert à étudier les moyennes des échantillons tirés d'une population normale.

Pour la loi de F, Z est définie par :

$Z = \frac{X}{Y}$ défini à partir d'une variable aléatoire X', suivant une loi de χ^2 à ν_X degrés de liberté par :

$$X = \frac{X'}{\nu_X}$$

Y est défini de la même façon à partir de Y' suivant la loi du χ^2 à ν_Y degrés de liberté, par :

$$Y = \frac{Y'}{\nu_Y}$$

La loi de distribution de Z est une loi "bêta" particulière c'est-à-dire comportant deux paramètres ν_X et ν_Y.

Le test F est la valeur telle que $\{ Z \leq F \} = 0,05$. Ce test a été précédé par un test moins commode "z" de Fischer qui considérait le logarithme népérien de Z. Comme cela arrive souvent en statistique, le test F a complètement supplanté le test "z".

La variable "t" de Student est définie par la valeur d'une variable aléatoire T utilisée dans le test HOTELLING.

ν est un paramètre et caractérise le nombre de degrés de liberté de la variable t.

$$t_\nu = \frac{X}{Y}$$

- X est une variable aléatoire, normale, centrée et normée.

- Y une variable aléatoire qui est la racine carrée d'une variable aléatoire V^2 qui a une distribution de χ^2 à ν degrés de liberté, divisée par la quantité ν

$$Y = \sqrt{\frac{V^2}{\nu}}$$

Cette variable aléatoire "t" peut être positive ou négative. Elle a pour espérance mathématique zéro.

En fait, la courbe représentative est à peu près celle de Laplace-Gauss.

Lorsque $\nu = 1$ on obtient pour t_1 la loi de Cauchy. Mais, dans ce cas, la racine carrée arithmétique de V^2, qui intervient dans la définition de t_ν, peut être remplacée par sa racine carrée algébrique, qui est, par définition du χ^2 une variable normale, centrée et normée.

- Nous remarquerons donc que dans ces conditions t_1 est le rapport de deux variables normales, de même variance, et de même espérance mathématique.

La loi de t_ν peut se calculer et nous la trouvons dans le chapitre premier.

- D'où le résultat :

Le quotient de deux variables normales et normées et de même variance ayant la même moyenne, est distribué suivant la loi de Cauchy.

<u>Cas du quotient de deux variables normales</u> :

Ces variables sont d'espérance mathématique $\begin{cases} \mu_X \\ \mu_Y \end{cases}$

et de variance σ_X^2 et σ_Y^2, quantités qui ne sont pas nécessairement égales.

On voit combien cette généralisation est restrictive puisque on s'intéressera à des populations normales.

F et t sont liées à des populations normales à cause du χ^2 malgré une généralité en apparence plus grande.

En outre, dans la définition de t, nous pouvons considérer que X est la racine carrée d'une variable distribuée comme un χ^2 avec 1 degré de liberté et nous retrouvons un résultat classique lorsque $\nu = 1$. t_1 est la racine carrée

de la variable qui correspond à F. On peut affirmer que l'utilisation du test de Student dans l'analyse de variance, donne des résultats identiques pour la comparaison des moyennes de deux échantillons.

Pour en revenir à $Z = \frac{X}{Y}$, cette fonction a été étudiée par GEARY en 1930.

Nous allons traiter tout d'abord de la loi de distribution :

Nous considérerons des variables normales :

espérance mathématique de $X = \mu_X$

espérance mathématique de $Y = \mu_Y$

Enfin, la covariance de X et Y, C(X, Y):

$$C(X, Y) = , \rho . \sigma X . \sigma Y$$

Pour simplifier l'écriture, nous étudierons la loi de la variable

$$A = \frac{\frac{X}{\sigma X}}{\frac{Y}{\sigma Y}} \rightarrow A = Z \times \frac{\sigma y}{\sigma X}$$

de valeur numérique a, ce qui revient à utiliser des variables normales et normées. Autrement dit, par analogie nous poserons :

$$\alpha = \frac{\frac{\mu X}{\sigma X}}{\frac{\mu Y}{\sigma Y}}$$

I) Si $\mu X = \mu Y = 0$.

On démontre que l'expression $\frac{a}{\sqrt{1 - \rho^2}}$ a une distribution de Cauchy ce qui n'a rien d'étonnant lorsqu'on songe que la variable t de Student d'ordre 1 possède cette distribution et que sa définition est exactement celle du rapport.

II) Si $\mu Y \neq 0$

$$u = \frac{\left|\frac{EY}{\sigma Y}\right| (a - \alpha)}{\sqrt{1 - 2a\rho + a^2}}$$

$$p(a)\, d(a) = \pm \left[y(y) + \varepsilon(u) \right] du$$

$$du = \frac{\left|\frac{EY}{\sigma Y}\right| \; a(\alpha - \rho) + 1 - \rho\alpha \; . \; da}{\left[1 - 2a\rho + a^2 \right]^{3/2}}$$

y(u) et $\varepsilon(u)$ sont des quantités assez compliquées. BASTIEN ne trouve que des résultats simples et intéressants que lorsque $\frac{\mu Y}{\sigma Y}$ est largement supérieur à 3.

Dans ces conditions le rapport $\frac{x}{y} = z$ peut être envisagé du point de vue de sa distribution.

Ce rapport a une fonction de distribution qui se réduit à la valeur (simple ?)

$$p_Z(z) = \frac{\mu X}{\mu Y} + u\sqrt{(\frac{\mu X}{\mu Y})^2 + (\frac{\sigma Y^2}{\mu Y^2} - 2\rho\frac{\sigma X\,\sigma Y}{\mu X\,\mu Y} + \frac{\sigma X^2}{\mu X^2})}$$

où u est la variable normale centrée réduite. Nous nous contenterons de ce résultat qui montre que Z a une distribution normale car c'est une fonction linéaire de u. Bien entendu, on suppose que le dénominateur ne saurait prendre la valeur 0 ou en d'autres termes, devenir négatif, ce qui est le cas des variables Biologiques.

BASTIEN précise dans son article que "l'étude du rapport de 2 variables normales , quant à la loi de ce rapport, n'a qu'un intérêt théorique car les deux variables ne seront qu'approximativement normales, et l'on ignore généralement la valeur des deux paramètres dont dépend cette loi."

Il s'agit d'une opinion très pessimiste et dans le courant de notre travail, nous allons chercher si nous ne pouvons pas ajuster les résultats expérimentaux des différents quotients aux valeurs de la loi normale, affectée des coefficients de BASTIEN.

Nous remarquerons que cette approximation est centrée autour du terme correspondant au développement de TAYLOR. Dans son article très intéressant GEARY donne l'exemple suivant emprunté à l'épidémiologie de la tuberculose. Ceci montre d'ailleurs que cette distribution du rapport a été recherchée pour des besoins médicaux à l'origine.

Donnons un résumé de cet exemple :

"Dans l'Etat libre d'Irlande la mortalité pour toutes les formes de tuberculose est de :

259,0 pour 10 000 en 1905

156,0 pour 10 000 en 1925

Nous pouvons constater que dans une localité bien déterminée, CASTLETOWN dont la population moyenne est de 10 000 habitants, il y a un accroissement de 5 %. "Est-ce que cet accroissement diffère de façon significative de l'abaissement de 40 pour cent relatif à l'état libre ?"

La façon dont le problème est formulé montre bien que cette technique est probablement largement applicable. En effet, on travaille sur des taux, et dans les publications médicales on ne connait pas le nombre réel de cas et l'effectif

de la population, mais simplement le taux (tant pour mille). Supposant que le taux de mortalité pour la tuberculose a été calculé en 1905 sur 10 000 personnes prises au hasard, dans celles de l'état libre, et que celui de 1925 a été calculé sur 10 000 personnes correspondant aux précédents (*), nous pouvons concevoir que l'échantillon a été sélectionné par maison ou par pièce d'habitation. Si le rapport de ces taux ne peut pas posséder une valeur supérieure à 1,05 (valeur du rapport pour CASTLETOWN), ou une valeur supérieure de façon improbable, Il n'y a pas de différence statistiquement significative entre le rapport de "CASTLETOWN et celui de l'Etat libre". Il est nécessaire, selon GEARY, d'introduire ce concept de populations correspondantes, à cause du coefficient de correlation entre le taux dans les mêmes localités pendant un certain laps de temps.

En fait le coefficient de correlation entre la mortalité pour la tuberculose dans un Comté pendant les intervalles de temps 1901.... 1910 et 1924... 1926, a été trouvé égal à + 0,8.

Il faut noter, ce que GEARY ne dit pas, que X et Y sont des taux de mortalité mais que l'on peut supposer leur distribution dans les divers états du monde et que cette distribution est normale.

Comme les valeurs numériques sont petites, cette distribution suit une loi de Poisson.

$$\left.\begin{aligned} \sigma X^2 &= \mu X \\ \sigma Y^2 &= \mu Y \end{aligned}\right) \text{ résultat qu'il étudie sans vergogne,}$$

de même que l'évaluation de ρ.

En somme, l'application du résultat de GEARY, aux taux de mortalité, nécessitera une discussion serrée de ces taux de mortalité et surtout de leur corrélation, notion que nous supposons comme des épidémiologistes.

Quoi qu'il en soit, dans l'équation qui donne u, si l'on remplace z par 1,05 (valeur pour CASTLETOWN), on trouve u = 3,57, ce qui est statistiquement significatif puisque u est distribué normalement, avec la variance unité, ce qui montre que la probabilité est bien inférieure à 0,01 de rencontrer par hasard, dans une distribution normée centrée, une valeur aussi grande.

Par suite, on peut conclure avec GEARY que CASTLETOWN se comporte d'une manière différente que le reste de l'Etat libre d'Irlande.

Dans cet exemple, X est le taux de mortalité en 1905 et μX vaut 259 tandis que Y est le taux de mortalité en 1925, μY vaut 156,0 pour 10 000 habitants

---

(*) Un échantillon apparié en quelque sorte.

Dans son article, NICKOLSON développe les résultats d'un autre exemple médical emprunté à l'obstétrique. "Certaines difficultés de l'accouchement sont dues aux disproportions entre la taille de la tête foetale et la taille des détroits qu'elle a à traverser.

Cette difficulté nécessite une césarienne dans environ 1% des cas (pour Guy's hospital), en 1937. La discussion purement obstétricale des données nous entrainerait trop loin, dans le cadre de ce chapitre, et conduit aux résultats suivants :

	Tête foetale	Surface du détroit supérieur
Moyenne en cm^2	65,8	121
Ecart quadratique moyen en cm^2	5,8	12,9

Ces variables sont distribuées à peu près normalement et nous pouvons les supposer indépendantes.

On applique alors la formule précédente à l'envers c'est-à-dire que l'on détermine la valeur de ce qui se trouve à l'intérieur d'une probabilité de 1% qui est celle des césariennes et l'on trouve la valeur de z qui correspond à cette probabilité 1 % maximale. On trouve alors un rapport des deux surfaces $z = 76,4$ %. Ce qui est intéressant est la valeur des paramètres qui entrent dans la formule et nous verrons dans les chapitres suivants l'usage que l'on peut en faire.

E T U D E E X P E R I M E N T A L E

****

INTRODUCTION

Pour cette partie pratique, il nous fallait trouver deux variables tout d'abord distribuées normalement, et dont le rapport ait un sens au point de vue Biologique.

Nos recherches nous ont montré que dans le cadre des statistiques appliquées à la Biologie, bien peu de variables sont normalement distribuées.

Tout d'abord nous avions pensé orienter notre étude sur le métabolisme basal. Mais ce dernier fait intervenir la surface des individus exprimée en mètres carrés. Cette surface est déterminée au moyen de tableaux en fonction de la taille et du poids des individus, et elle n'est pas calculée de façon très précise.

Aussi avons-nous orienté nos recherches vers d'autres rapports.

Tout d'abord le rapport $\frac{\text{gamma-globulines}}{\text{sérum albumines}}$ (les données expérimentales nous ont été données par l'Hôpital Robert PIQUE).

D'autre part, le rapport $\frac{E}{T}$ de deux comptages radio-actifs. Les valeurs expérimentales nous ont été fournies par Monsieur le Professeur BLANQUET, qui nous a d'ailleurs donné toutes les explications concernant le "test Tétrasorb" où interviennent les deux variables E et T.

Une autre difficulté se présente : les deux variables retenues ne sont jamais pleinement indépendantes. Or cette indépendance est le fondement des résultats de GEARY. Cela va nous être l'occasion de voir si une liaison les affecte, ou, comme on dit, si la méthode de GEARY est robuste.

CHAPITRE PREMIER

ETUDE DU RAPPORT DE DEUX VARIABLES NORMALES ET LIEES ENTRE ELLES. LES GAMMA-GLOBULINES ET LES SERUM ALBUMINES

(Les variables aléatoires sont normales et liées entre elles, pour les cent valeurs que nous avons considérées. Nous ne pouvons pas généraliser.)

ETUDE DU RAPPORT GAMMA-GLOBULINES

SERUM ALBUMINES

Un protidogramme révèle 5 fractions.

1 - sérum albumines 2 - α_1 globulines 3 - α_2 globulines

4 - β globulines 5 - γ globulines.

Les valeurs de ces fractions s'expriment en pourcentage.

JAULMES a donné les pourcentages suivants :

Albumines : 53,3 à 63,4 %
Globulines γ : 11 à 15,4 %

Notre étude a porté sur cent malades de l'Hôpital ROBERT-PIQUÉ

Pour les albumines, nous avons trouvé un pourcentage moyen de 56,42 avec une variance de 35,83 et un écart type de 5,98 %. Le coefficient de variation est donc 0,106.

Pour les gamma-globulines, nous avons trouvé un pourcentage moyen de 18,56 % avec une variance de 19,09 % et un écart type de 4,36 %.

Le coefficient de variation est donc 0,235

Ce sont ces taux individuels d'albumine et de globuline qui sont les variables aléatoires.

I - Etude de la loi de distribution des sérum albumines

Nous avons divisé notre population en 11 classes allant de 39 % à 70 %

Nous avons ainsi dressé le tableau

Pourcentage de sérum-albumines	Nombre de malades
39..... 42 %	1
42..... 45 %	3
45..... 48 %	5
48..... 51 %	8
51..... 54 %	20
54..... 57 %	19
57..... 60 %	22
60..... 63 %	9
63..... 66 %	7
66..... 69 %	4
69..... 72 %	2

Nous avons dressé un histogramme, ainsi que le tracé de la droite de HENRY correspondant à cette population.

Puis nous avons ajusté notre échantillon de moyenne 56,42 et d'écart type 5,98, à une loi normale. Nous avons pu ainsi comparer l'effectif théorique et l'effectif réel et faire le calcul du χ^2.

TABLEAU DE CONTINGENCE

Limites de classe 0	Classes	Effectif théorique Ti	Effectif Réel Ré
39 %	/	0,2	0
42 %	1	0,6	1
45 %	2	2,5	3
48 %	3	5	5
51 %	4	9,5	8
54 %	5	16,2	20
57 %	6	19,5	19
60 %	7	18,5	22
63 %	8	14,5	9
66 %	9	8	7
69 %	10	3,7	4
72 %	11	1,3	2
100 %	/	0,5	0

L'écart quadratique réduit est pour chaque classe :

$\frac{(Ri - Ti)^2}{Ti}$ Ri : effectif réel

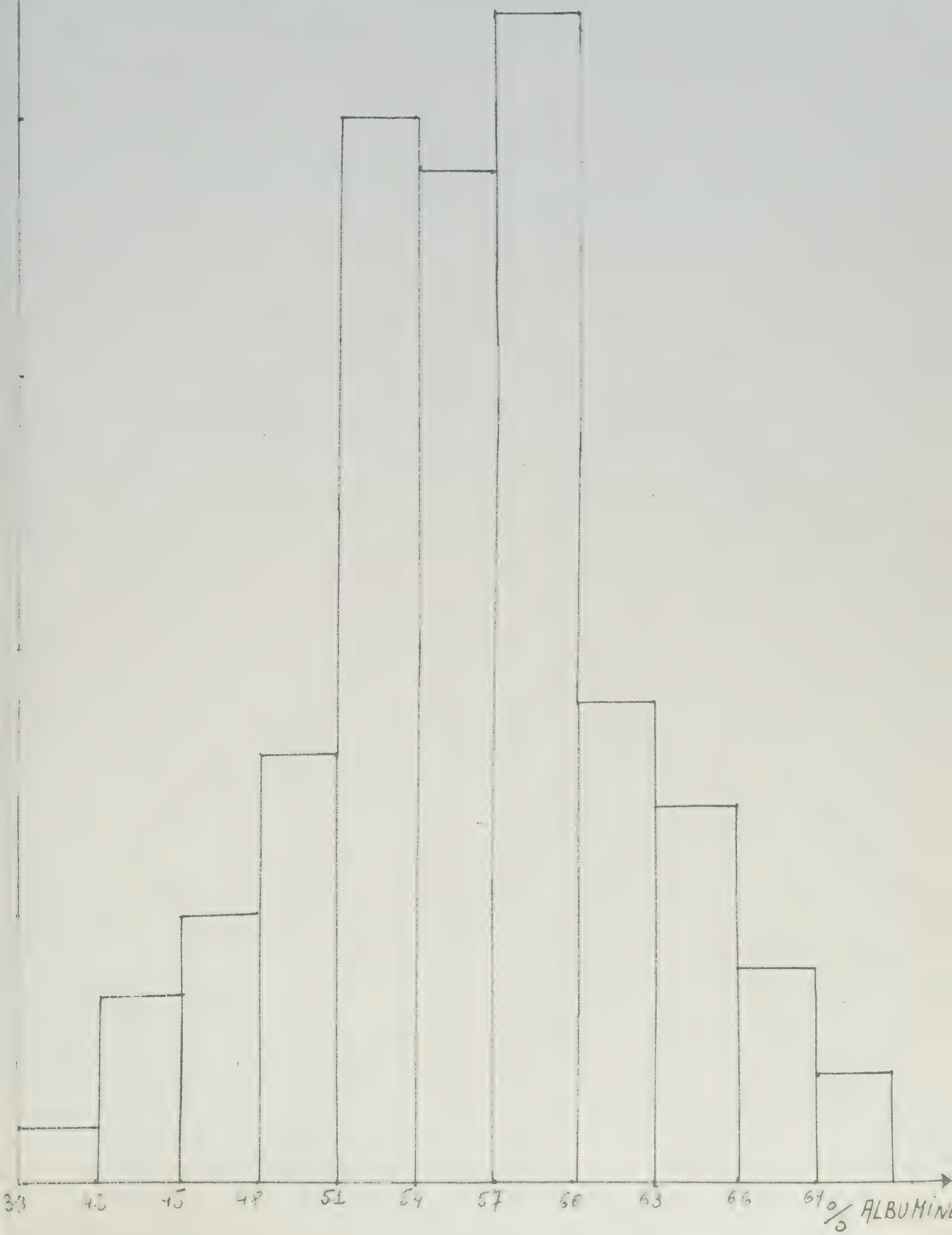
Ti : effectif théorique

χ^2 est la somme des écarts quadratiques réduits

$\chi^2 = \sum_{t=1}^{n} \frac{(Ri - Ti)^2}{Ti} = 5$ (non significatif) pour $\nu = 6$

Nombre de malades

Histogramme ①



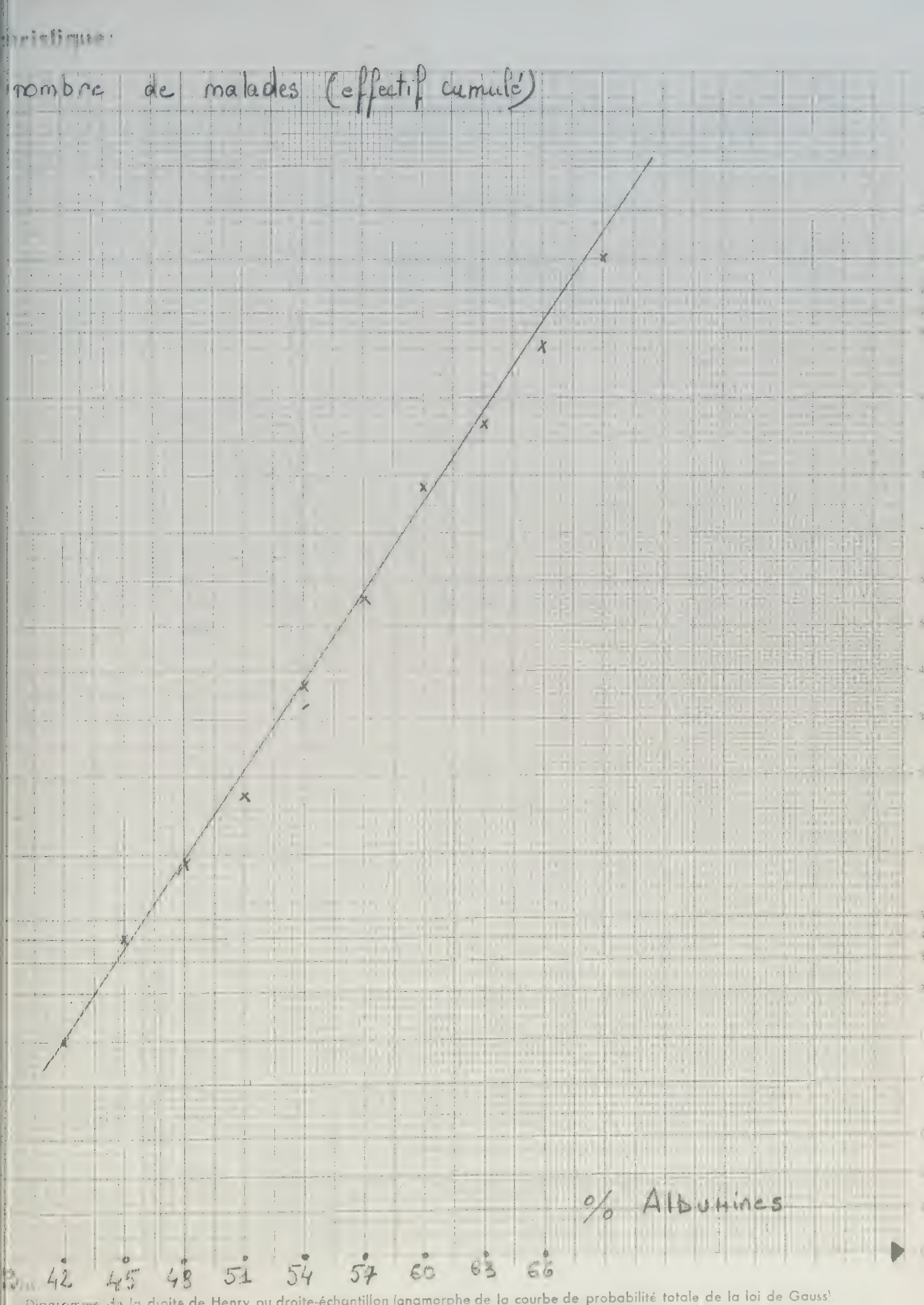


Diagramme de la droite de Henry ou droite-échantillon (anamorphe de la courbe de probabilité totale de la loi de Gauss).

Ceci montre que nous pouvons assimiler la loi de distritution des sérum-albumines à une loi normale, dans le cas de nos cent valeurs.

II - Etude de la loi de Distribution des Gamma-globulines

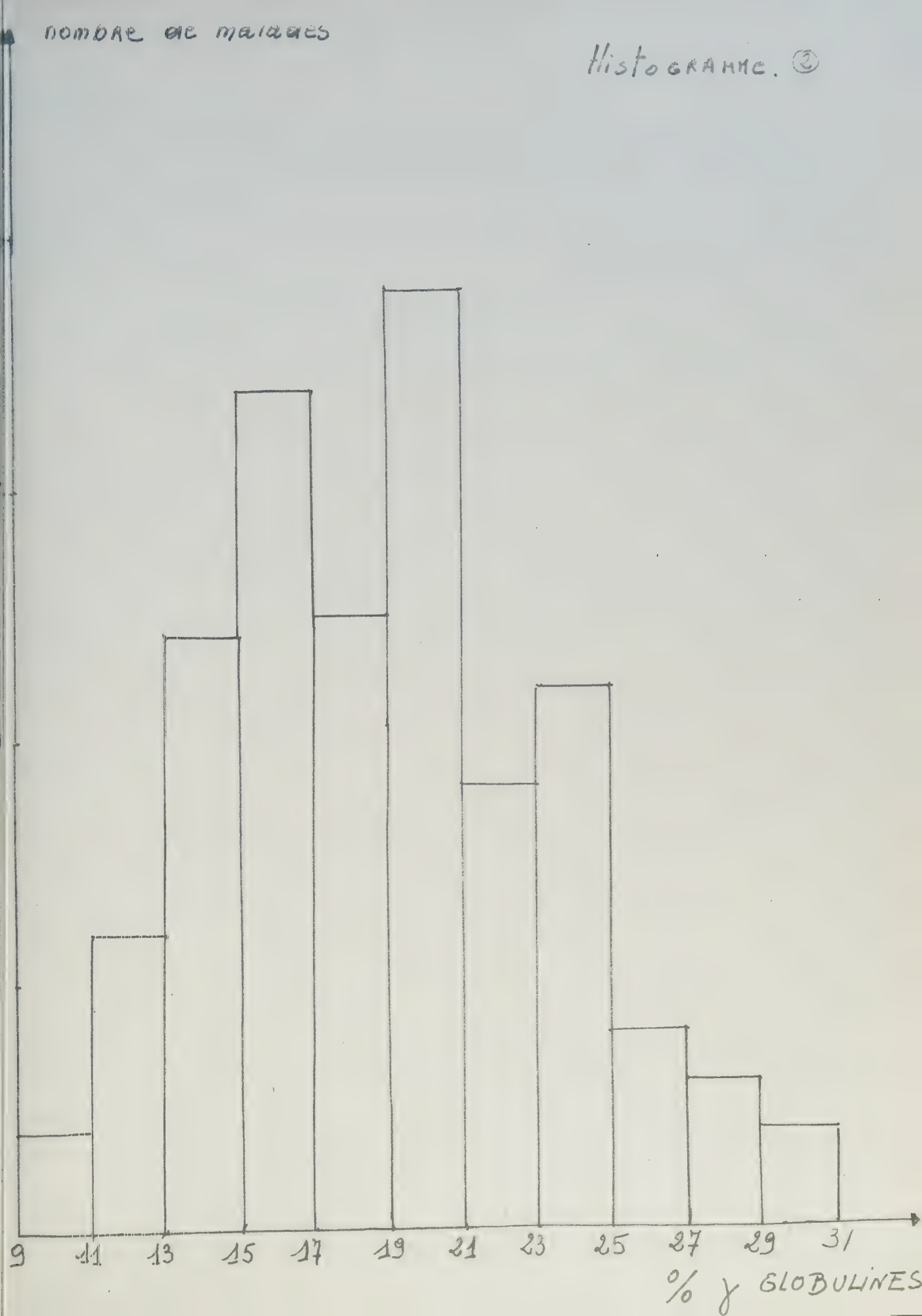
Nous adoptons le même plan d'étude, pour cette distribution. Nous avons pris 100 valeurs de gamma-globulines correspondant aux cent valeurs des albumines. Après avoir divisé la population des gamma-globulines en 11 classes nous avons dressé le tableau :

GAMMA-GLOBULINES	NOMBRE DE MALADES
9 %..... 11 %	2
11 % ..... 13 %	6
13 %..... 15 %	12
15 % ..... 17 %	17
17 %..... 19 %	15
19 %..... 21 %	19
21 % ..... 23 %	9
23 %..... 25 %	11
25 %..... 27 %	4
27 % ..... 29 %	3
29 %..... 31 %	2

Ensuite, nous avons tracé l'histogramme ainsi que la droite de Henry.

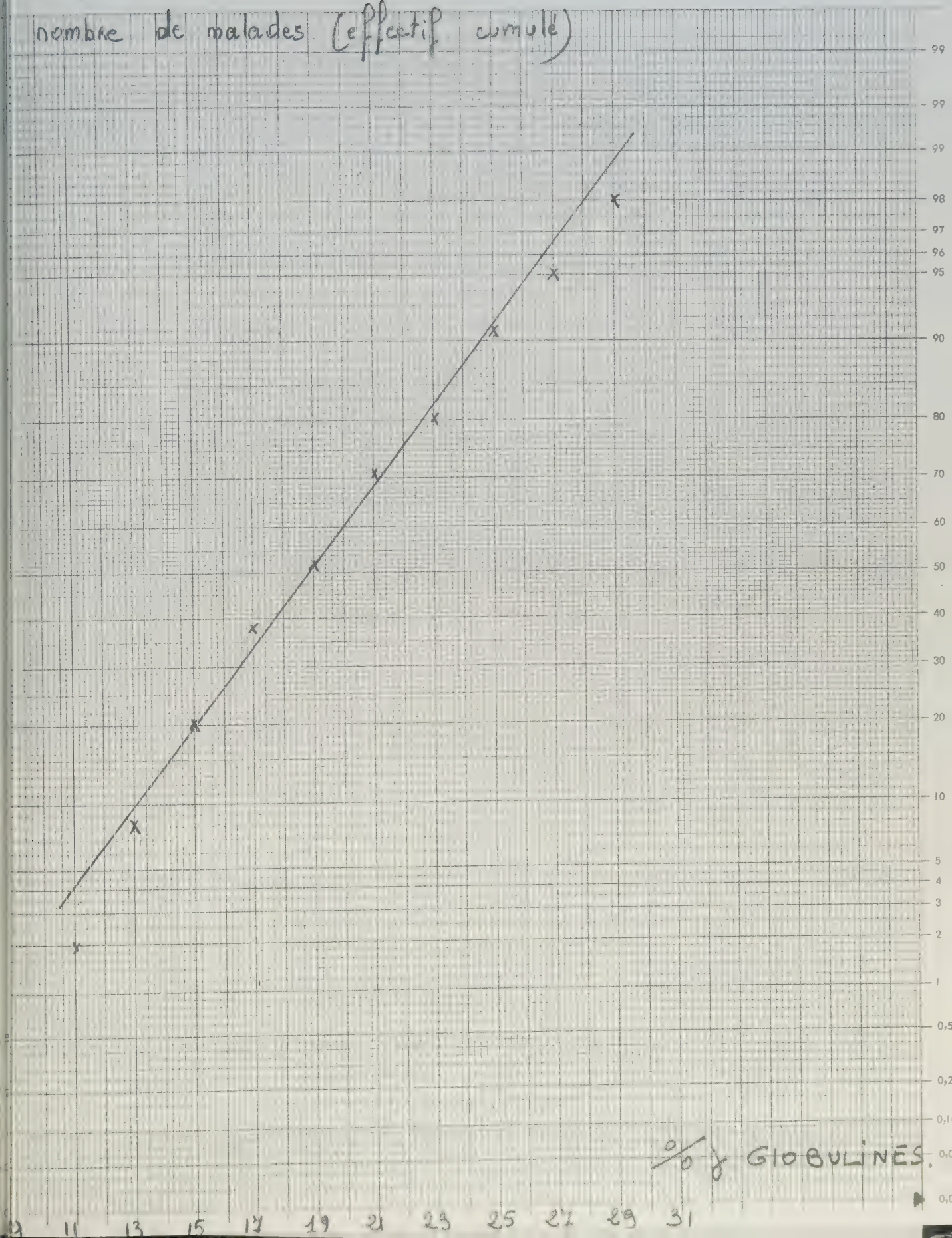
Nous avons ajusté la loi de distribution de la population à une loi normale.

Histogramme. ②



IIIT

éristique:



LIMITES DE CLASSES	CLASSES	EFFECTIF THEORIQUE	EFFECTIF REEL
0	/	1,8	0
9			
	1	2,6	2
11			
	2	5,5	6
13			
	3	11	12
15			
	4	15	17
17			
	5	18	15
19			
	6	18	19
21			
	7	13	9
23			
	8	8	11
25			
	9	4	4
27			
	10	2	3
29			
	11	0,9	2
31			
	/	0,2	0
100			

avec $\nu = 6$

La calcul du χ^2 nous donne 4,06 (non significatif). Nous pouvons donc considérer la loi de distribution des gamma-globulines comme normale, pour nos cent valeurs.

III - Etude de la loi de distribution du rapport : $\frac{\text{Gamma-globulines}}{\text{Sérum-albumines}}$

Les lois de distribution des gamma-globulines d'une part, des sérum-albumines d'autre part, étant normales, nous nous sommes demandés si, expérimentalement, la loi de distribution du rapport $\frac{\text{gamma-globulines}}{\text{sérum-albumines}}$ était normale.

Nous avons commencé par calculer le rapport $\frac{\text{gamma-globulines}}{\text{sérum-albumines}}$, pour les cent malades.

Nous avons donc obtenu une nouvelle population. Les nombres exprimant ce rapport étant naturellement inférieurs à l'unité, ont été multipliés par cent si bien que le rapport représente un pourcentage.

Cette population présente une moyenne de 33,93 % avec une variance de 123,82 % et un écart type de 11,12 %.

Le coefficient de variation est de 0,327.

Nous avons divisé cette population encore en 11 classes allant de 15 % à 75 %.

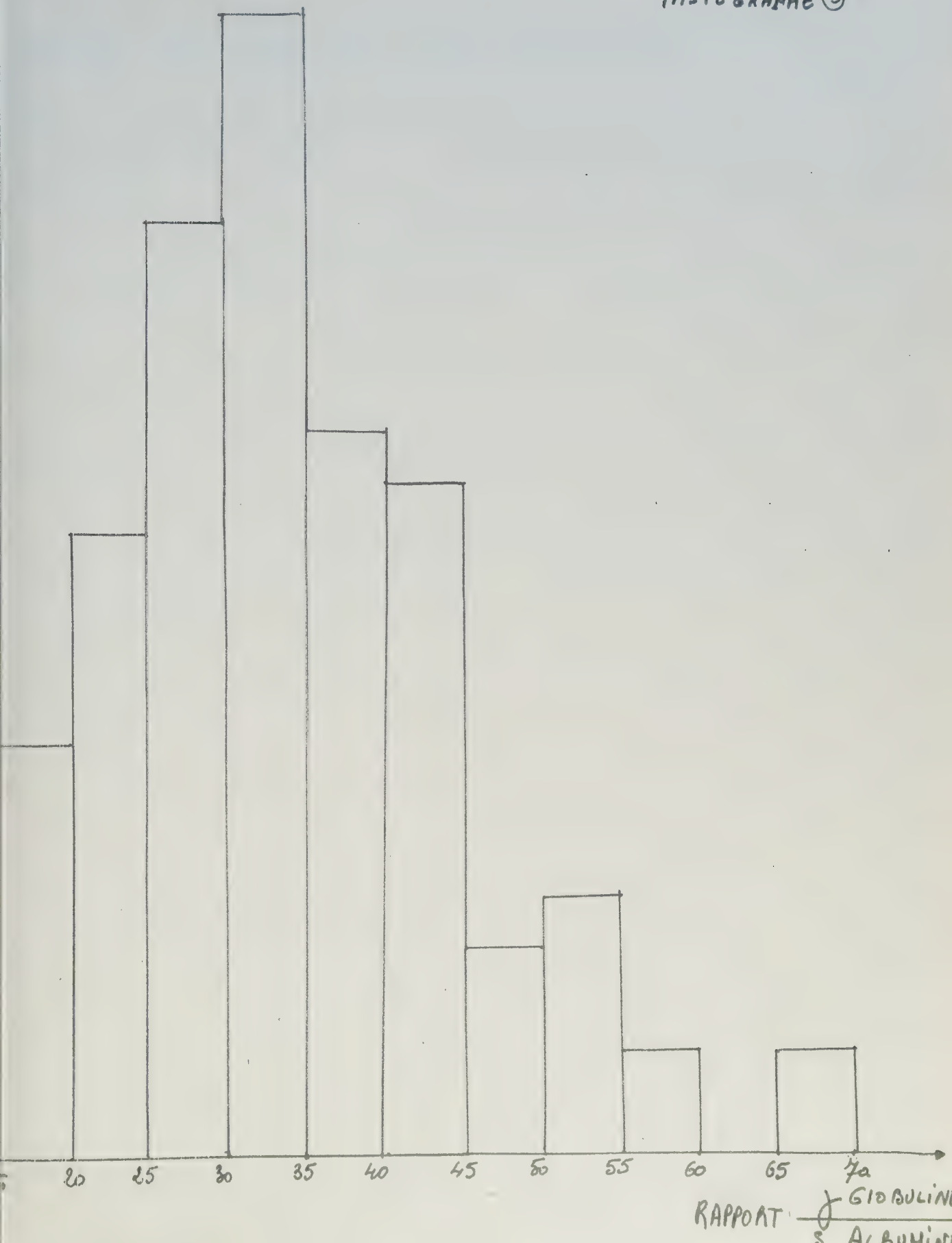
On a pu dresser le tableau suivant, tracer la droite de Henry, qui s'est révélée assez satisfaisante, et l'histogramme.

POURCENTAGES	NOMBRE DE MALADES
15 % ..... 20 %	8
20 % ..... 25 %	12
25 % ..... 30 %	18
30 % ..... 35 %	22
35 % ..... 40 %	14
40 % ..... 45 %	13
45 % ..... 50 %	4
50 % ..... 55 %	5
55 % ..... 60 %	2
60 % ..... 65 %	0
65 % ..... 70 %	2

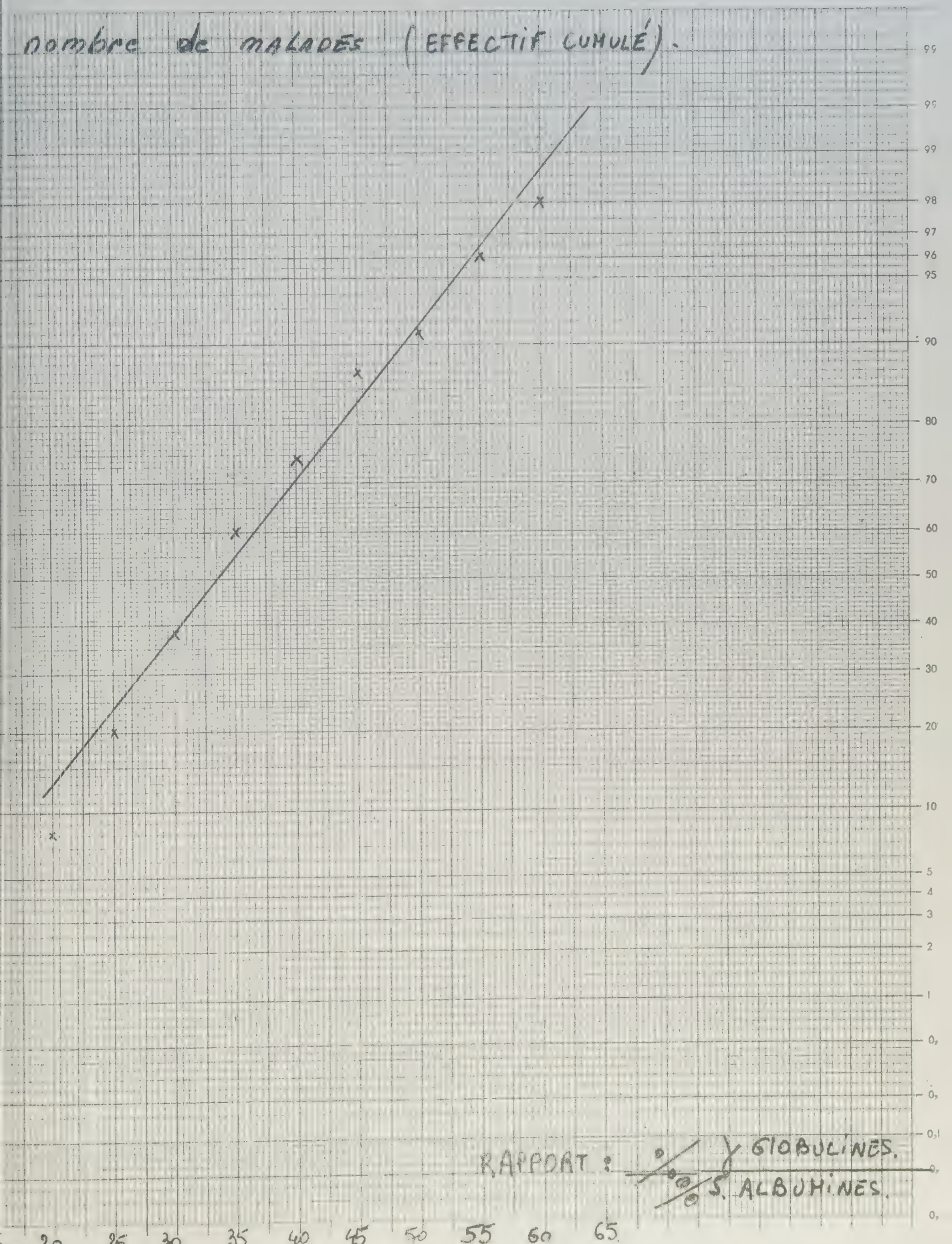
Enfin, nous avons vu que l'on pouvait ajuster la loi de distribution du rapport à une loi normale. En effet, le calcul du χ^2 nous a donné 5,91, ce qui n'est pas encore significatif vu que $\nu = 7$.

nombre de malades

HISTOGRAMME ③



RAPPORT $\frac{\text{GLOBULINES}}{\text{S. ALBUMINES}}$



715

Diagramme de Henry ou droite échantillon (anamorphose de la courbe de probabilité totale de la loi de Gauss).

TABLEAU DE CONTINGENCE

INTERVALLES DE CLASSES	CLASSES	EFFECTIF THEORIQUE	EFFECTIF REEL
0	/	4,5	0
15	1	7	8
20	2	11	12
25	3	15	18
30	4	17	22
35	5	16	14
40	6	13	13
45	7	8,5	4
50	8	5	5
55	9	2,5	2
60	10	0,8	0
65	11	0,15	2
70	/	0,1	0
100			

IV - Etude de la liaison entre sérum-albumines et gamma-globulines

Appelons x la valeur du pourcentage des sérum-albumines et y la valeur du pourcentage des gamma-globulines.

somme des x = 1856,1

somme des x^2 = 36 360,44

somme des y = 5 642,6

somme des y^2 = 321 972,96

Nous avons appareillé les valeurs de x et y pour un même individu et nous avons calculé l'expression $\sum_{0}^{100} x.y.$

Nous avons trouvé 103 117,85

Puis nous avons cherché l'équation de la droite de régression de y en x. Nous avons trouvé :

$Y = - 0,8460 X + 72,11$

Le calcul du coefficient de corrélation a donné le résultat suivant :

$$r = - 0,617$$

Ceci montre que les variables X et Y sont indiscutablement liées.

V - Applications de l'approximation de Geary

BASTIEN dit, dans son article, que l'on ignore généralement les deux paramètres dont dépend la loi du rapport de deux variables, et que les variables ne sont que sensiblement normales.

Dans notre travail, nous avons montré que ces variables étaient normales, certes sensiblement normales.

Mais nous avons pu calculer les paramètres dont dépend la loi du rapport de deux variables.

Nous avons trouvé expérimentalement :

moyenne 33,93 % soit 0,3 393

écart type 11,12 % soit 0, 1112.

Voyons quelle valeur peut-on attendre de l'approximation de Geary sur un exemple biologique.

1º) Nous avons tout d'abord cherché si l'approximation pourrait s'appliquer à notre travail.

Les calculs mathématiques montrent que cette approximation n'est valable que si le rapport $\frac{\text{moyenne des v. albumines}}{\text{écart type des v. albumines}}$ est supérieur à 3.

Calculons ce rapport, nous trouvons : 9,4

Donc, l'approximation de Geary est applicable à notre travail.

2º) Cherchons si l'approximation de Geary nous donne, quant à la loi de distribution du rapport, un écart type et une moyenne voisins de ceux trouvés expérimentalement.

Appelons :

μY : la moyenne des valeurs de la variable albumine

μX : la moyenne des valeurs de la variable gamma-globuline

Nous avons :

μY : 56,42 % soit 0,5642

μX : 18,56 % soit 0,1856

σY : l'écart type correspondant aux albumines

σX : l'écart type correspondant aux gamma-globulines.

D'après Geary, la loi de distribution du rapport $\frac{X}{Y}$ peut s'exprimer au moyen de l'approximation suivante :

$$\frac{x}{y} = \frac{\mu X}{\mu Y} + u \sqrt{(\frac{\mu X}{\mu Y})^2 + (\frac{\sigma Y^2}{\mu Y^2} - 2\rho \frac{\sigma_X \sigma_Y}{\mu X \mu Y} + \frac{\sigma X^2}{\mu X^2})}$$

$\sigma X = 4,36$

$\sigma Y = 5,98$

Nous trouvons donc :

$$\frac{x}{y} = 0,3289 + u \times 0,1027.$$

D'après Geary, la moyenne des valeurs de notre population $\frac{X}{Y}$ est 0,3289. Expérimentalement nous trouvons 0,3393.

L'écart type est 0,1027. Expérimentalement nous avons 0,112.

Geary a donc donné, comme le dit Bastien, une bonne approximation de la loi du rapport de deux variables, du moins, dans le cadre des statistiques médicales.

Certes il est difficile de trouver des variables normales, et nous étudierons dans le chapitre suivant le cas des variables non normales.

Auparavant, donnons un ou deux exemples d'application de l'approximation de Geary.

Dans la population des malades de l'Hôpital Robert PIQUÉ, nous avons un malade dont le rapport $\frac{\text{gamma-globulines}}{\text{sérum-albumines}}$ vaut 0,74.

Que faut-il en penser ?

$0,74 = 0,3289 + u \times 0,1027$ (u variable normée centrée)

$u \simeq 4$

Consultons la table :

Nous voyons que la probabilité de trouver un tel malade est très très faible.

- Inversement, posons $u = \pm 1$

$$\frac{x}{y} = 0,3289 \pm 0,102$$

Il apparaît que si le rapport $\frac{x}{y}$ est compris entre 0,2269 et 0,4309, il y a 68 chances sur cent, d'après la table, de trouver un tel rapport.

Nous avons calculé le nombre de malades, de cette manière donc, dont le rapport $\frac{x}{y}$ était compris entre 0,2269 et 0,4309. Nous en avons trouvé 68 puisque nous avions une population de cent malades.

Nous avons compté, expérimentalement, le nombre de malades dont le rapport était effectivement compris entre 0,2269 et 0,4309. Nous en avons trouvé 72.

Dans le cadre des variables normales, l'approximation de GEARY est donc largement applicable.

Voyons maintenant, le cas des variables non normales.

CHAPITRE SECOND

ETUDE DU RAPPORT $\frac{E}{T}$: $\frac{\text{Eponge}}{\text{Totalité}}$

E et T sont deux variables aléatoires non normales, et non liées, pour les mille valeurs que nous avons étudiées, ce sont des variables intervenant dans le test "Tétrasorb".

ETUDE DU RAPPORT $\frac{E}{T}$

(E et T sont deux variables intervenant dans le test "Tétrasorb")

Dans ce travail, nous avons utilisé les dornées expérimentales, qui nous ont été fournies par Monsieur le Professeur BLANQUET, et nous nous sommes reportés à la Thèse de Monsieur ROUAIS.

I - Le test Tétrasorb

Ce test est une méthode de détermination de la concentration totale de T_4 sérique, basée sur la détermination d'un rapport de deux variables aléatoires, d'où l'intérêt pour notre travail.

Sous l'appellation de Thyroxine Binding Protein, on désigne l'ensemble des protéïnes capables de fixer la thyroxine.

L'affinité de la thyroxine est maximale à l'égard de la T B G (inter α_1. α_2 globuline).

L'affinité de l'hormone est telle, que l'on a une liaison et la liaison hormone-protéïne, obéit à une réaction d'équilibre régie par la loi d'action de masse.

$$T_4 + TBG \rightleftharpoons TBG - T_4.$$

$$K = \frac{\left[T_4 - TBG\right]}{\left[T_4\right]\left[TBG\right]} \qquad K = 10^{10}.$$

La quantité d'hormone fixée par la protéïne porteuse, dépend non seulement de l'affinité de la thyroxine, pour cette protéïne, mais aussi du nombre de sites de cette dernière. En effet, les protéïnes porteuses possèdent une capacité vectrice limitée, celle-ci étant vraisemblablement fonction de leur propre concentration sérique. En conséquence, le taux de T_4 libre du sérum, paraît être en rapport avec la capacité de la liaison de la TBG, et, en définitive, avec sa concentration sérique.

Il existe une faible concentration de TBG dans le sérum. Les sites accepteurs sont facilement saturés par une petite quantité de T_4. Si l'on ajoute en même temps de la T_4 I^{125} comme traceur, la concentration de T_4 liée au protéïnes

va pouvoir être déterminée.

En effet, à mesure que l'on ajoute de la T_4 non marquée au système, la quantité Thyroxine I^{125} liée aux protéïnes va diminuer, puisque les Thyroxines marquées et non marquées entrent en compétition pour les mêmes sites. Lorsque l'équilibre est atteint, l'intervention d'une éponge-résine, fixant la T_4 libre permet la séparation de la T_4 liée à la TBG et de la T_4 libre.

En pratique :

On met en circulation la T_4 d'un sérum à déterminer avec un complexe (TBG - T_4*) contenu dans des seringues et tout préparé).

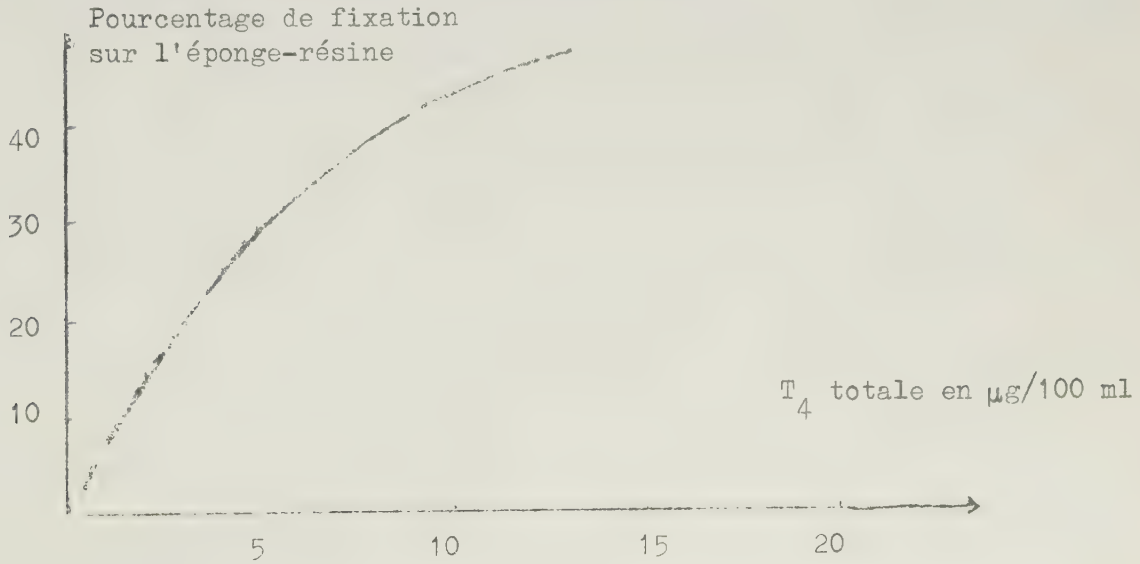
Donc la T_4 de l'échantillon entre en compétition avec la T_4* pour les sites de liaison de la TBG.

La Thyroxine non liée se fixe sur la résine échangeuse d'ion avec une vitesse proportionnelle à la concentration de T_4 du sérum.

On détermine la radio-activité initiale du mélange après 30 minutes d'incubation : c'est la variable T.(Totalité). Ensuite onlave l'éponge-résine. On compte la radio-activité de l'éponge qui correspond à la fraction de Thyroxine libre : c'est la variable E ("Eponge"). Pour connaître le pourcentage de fixation, on calcule le rapport suivant :

$$\frac{\text{imp./min. correspondant à la radio-activité de l'éponge}}{\text{imp./min. initiale}} \times 100$$

Pour obtenir la concentration du sérum en µg/ml de T_4 totale, on se reporte à la courbe-étalon déterminée préalablement.



II - Etude de la distribution de la variable"Eponge".

Nous avons pris une population d'effectif N = 1000.

Nous avons calculé la moyenne, la variance, l'écart type, le coefficient de variation. La loi de distribution de la variable Eponge, ne nous est pas apparue comme une loi normale.

Moyenne : 7 965,563

Variance : 8 860 685,4

Ecart type : 2 376,69

Coefficient de variation : 0,373

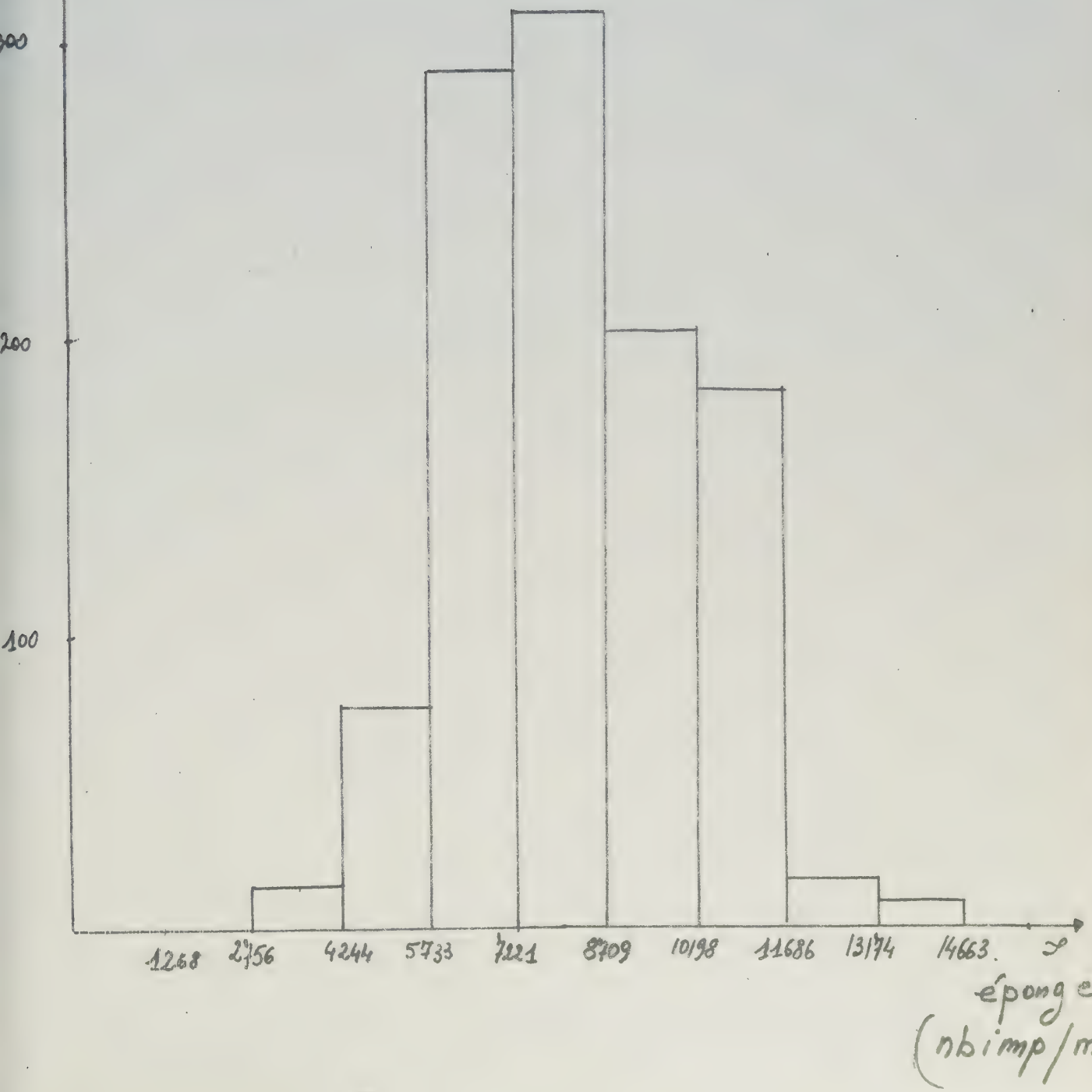
Nous avons divisé ensuite, la population en 11 classes ayant comme limites 1268 et 14 663,11 (0 et ∞ étant les limites extrêmes).

Nous avons essayé d'ajuster la loi de distribution à une loi normale.

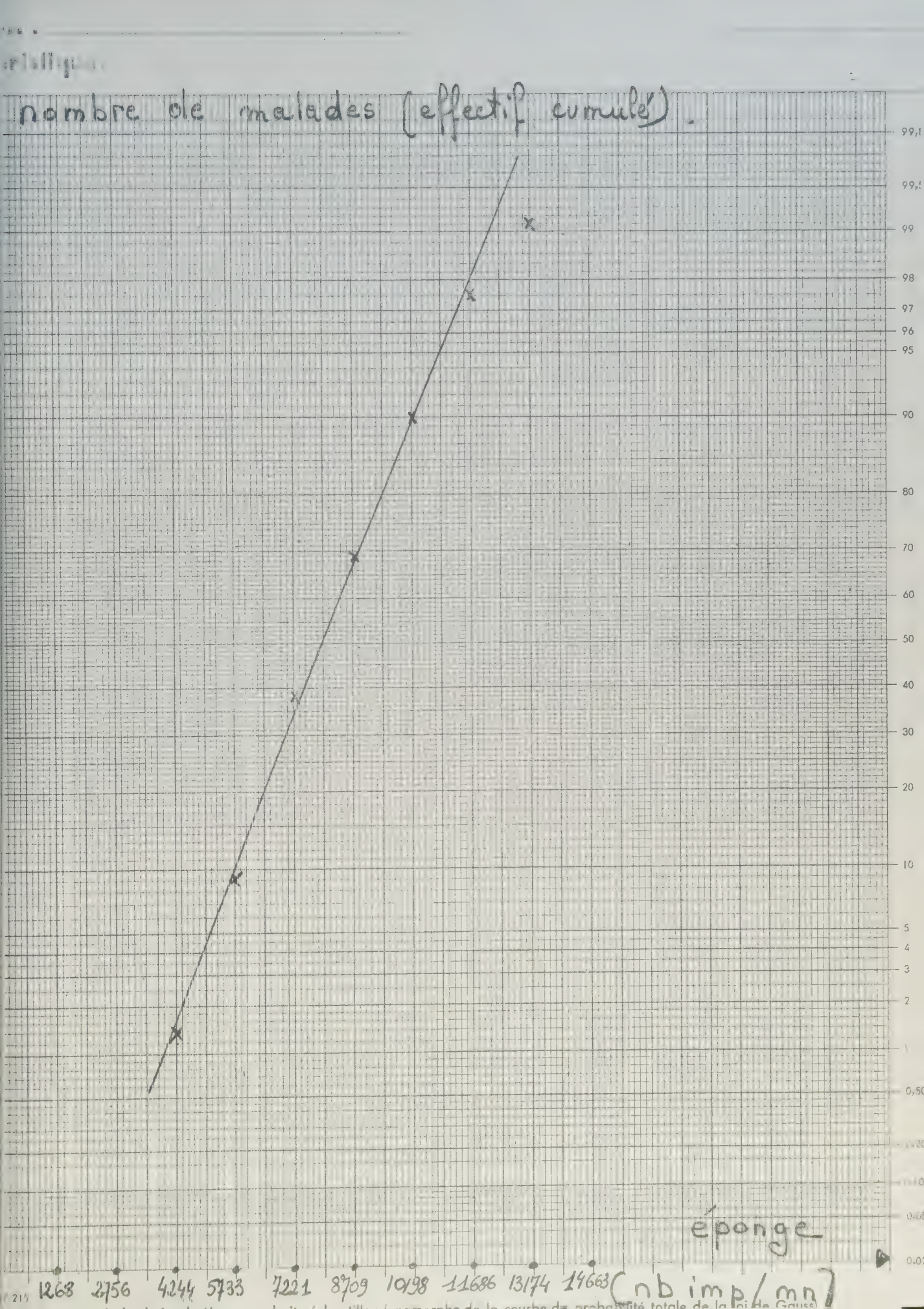
CLASSE	LIMITE 0	EFFECTIF THEORIQUE Ti	EFFECTIF EXPERIMENTAL Ri
1		12,22	0
	1 268		
2		27,83	0
	2 756,3		
3		65,59	14
	4 244,7		
4		120,9	75
	5 733,04		
5		174,66	230
	7 221,39		
6		197,42	310
	8 709,73		
7		174,66	203
	10 198,08		
8		120,98	203
	11 686,4		
9		65,59	16
	13 174,77		
10		27,83	8
	14 663,11		
11		12,22	1
	+ ∞		

nombre de malades. (effectif non cumulé)

Histogramme ④



nombre de malades (effectif cumulé).



Nous avons tracé, ensuite, la droite de Henry ainsi que l'histogramme correspondant à cette population.

Nous avons fait ensuite un calcul de χ^2

$$\chi^2 = \sum_{0}^{11} \frac{(Ri - Ti)^2}{Ti} = 680 \text{ !!} \qquad \nu = 10$$

Ce χ^2 est très significatif. Nous ne pouvons pas considérer que la population "éponge" est distribuée suivant une loi de Laplace-Gauss.

III - Etude de la distribution de la variable "Totalité".

Là encore, notre étude nous a montré que la variable "Totalité" n'est pas distribuée suivant une loi de Laplace-Gauss.

Nous avons pris les mille valeurs de la variable"totalité" correspondant naturellement aux mille valeurs de la variable "éponge".

Nous avons calculé, toujours, moyenne, variance, écart type. Nous avons trouvé :

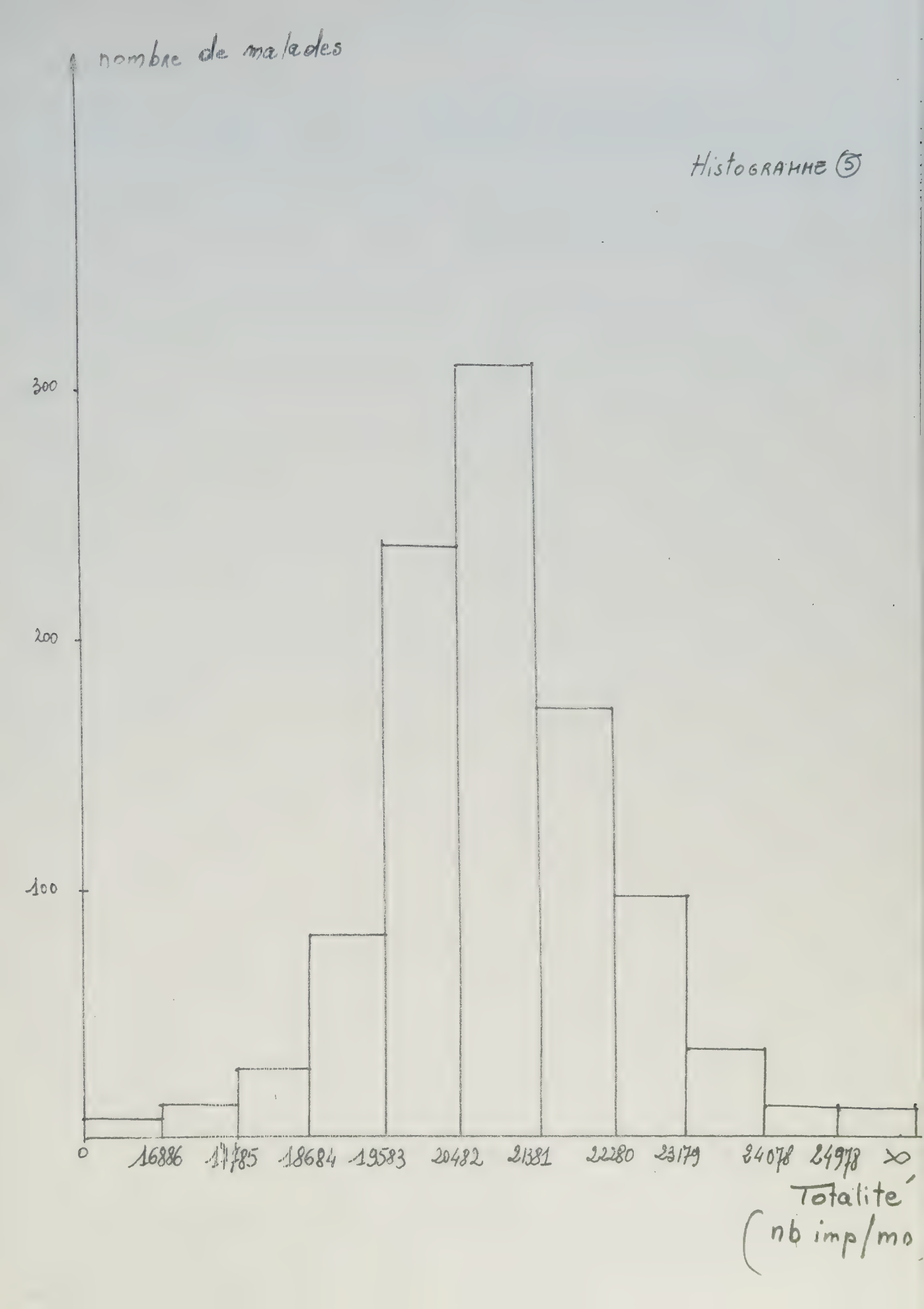
moyenne : 20 931,589

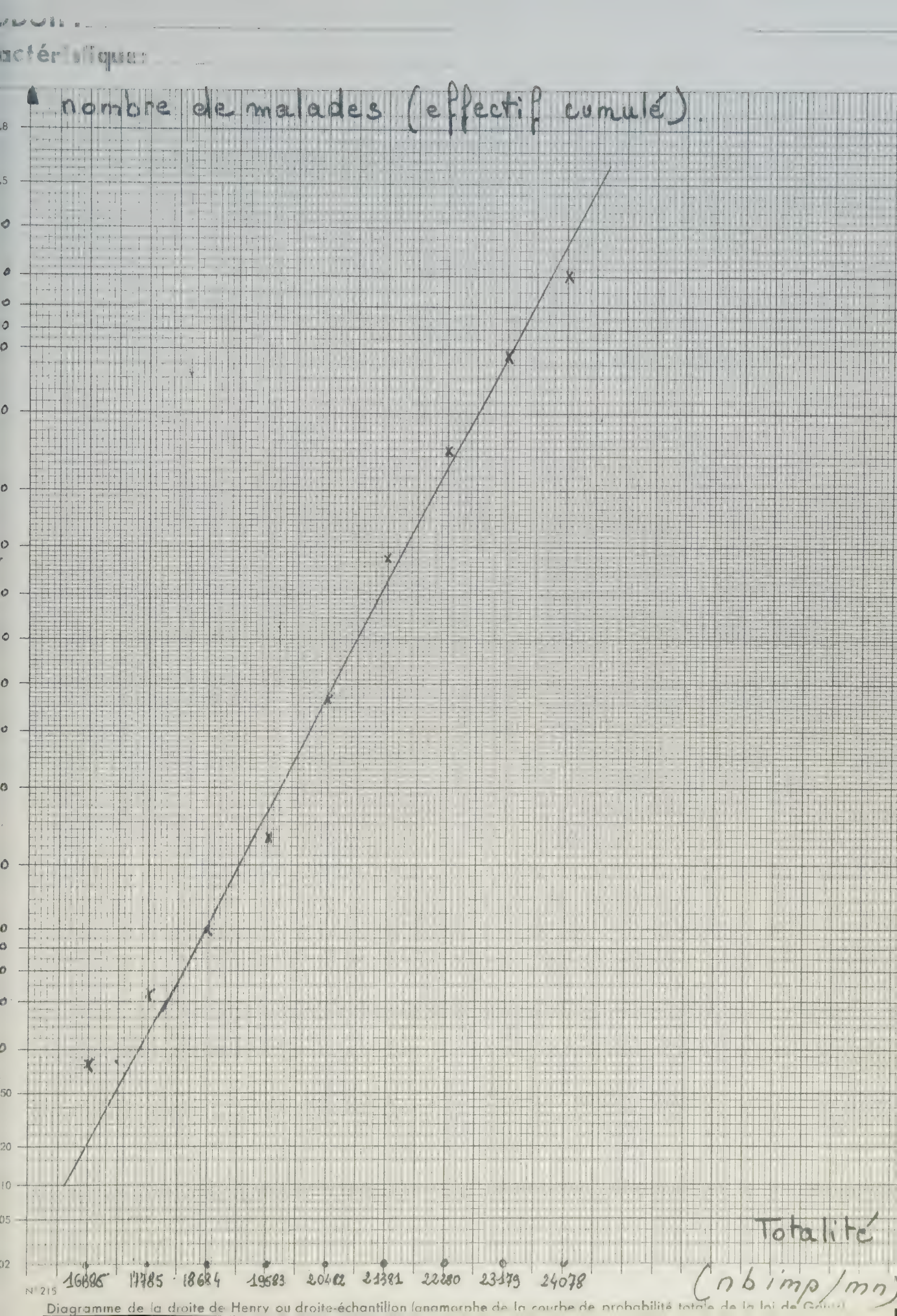
Variance : 3 233 565

Ecart type : 1 798,21

Coefficient de variation : 0,08590

Nous avons divisé notre population en 11 classes allant de 16 885,61 à 24 977,56 et nous avons essayé de faire un ajustement à une loi de Gauss.


nombre de malades
Histogramme (5)
300
200
100
0
16886
17785
18684
19583
20482
21381
22280
23179
24078
24978
∞
Totalité
(nb imp/mo



N° 215

Diagramme de la droite de Henry ou droite-échantillon (anamorphe de la courbe de probabilité totale de la loi de Gauss)

CLASSES	LIMITES	EFFECTIF THEORIQUE	EFFECTIF REEL
	0		
1		12, 224	8
	16 885,61		
2		27,835	14
	17 784,71		
3		65,591	28
	18 683,82		
4		120,38	81
	19 582,93		
5		174,66	238
	20 482,03		
6		120,98	305
	21 381,14		
7		65,591	179
	22 280,247		
8		120,98	94
	23 179,35		
9		65,59	33
	24 078,45		
10		27,83	10
	24 977,56		
11		12,224	10
	+ ∞		

Le calcul du χ^2 nous a donné :

$\chi^2 = 159$!! $\nu = 10$.

Ce χ^2 est très significatif.

La loi de distribution des variables "Totalité" ne peut pas être considérée comme normale.

IV - Etude du rapport des variables aléatoires $\frac{\text{Eponge}}{\text{Totalité}}$

Nous avons apparié les valeurs des variables éponge et totalité, pour un même malade, et nous avons étudié les valeurs des mille variables aléatoires $\frac{E}{T}$.

En consultant les chiffres, nous avons remarqué qu'à l'intérieur de cette population d'effectif N = 1000, il y avait des lots où l'on pouvait considérer la loi de distribution des variables $\frac{E}{T}$ comme normale, d'autres lots où elle n'était pas du tout normale. Ainsi nous avons divisé cette population en 10 classes d'effectif N = 100, naturellement.

Ainsi, nous avons calculé la moyenne, la variance, l'écart type. Puis nous avons fait un calcul de χ^2, et dressé un histogramme. (Nous avons ajusté à une courbe normale, un échantillon connu par sa moyenne et son écart type).

Ensuite, nous avons étudié la population d'effectif N = 1000 en bloc.

I - ETUDE DU LOT nº 1

(Les résultats sont exprimés en pourcentage)

N = 100 ; Moyenne = 35,9 %

Ecart type = 8,24

Coefficient de variation = 0,229

SOUS-CLASSES	LIMITES	EFFECTIF THEORIQUE	EFFECTIF EXPERIMENTAL
	0		0
1		1,2 224	
	17,36		
2		2,7835	5
	21,48		
3		6,5591	5
	25,6		
4		12,098	9
	29,72		
5		17,466	24
	33,84		
6		19,742	23
	37,96		
7		17,466	11
	42,08		
8		12,098	9
	46,2		
9		6,5591	9
	50,32		
10		2,7835	3
	54,44		
11		1,2224	2
	100		

Nous avons ensuite dressé l'histogramme nº I, puis nous avons fait le calcul du χ^2 :

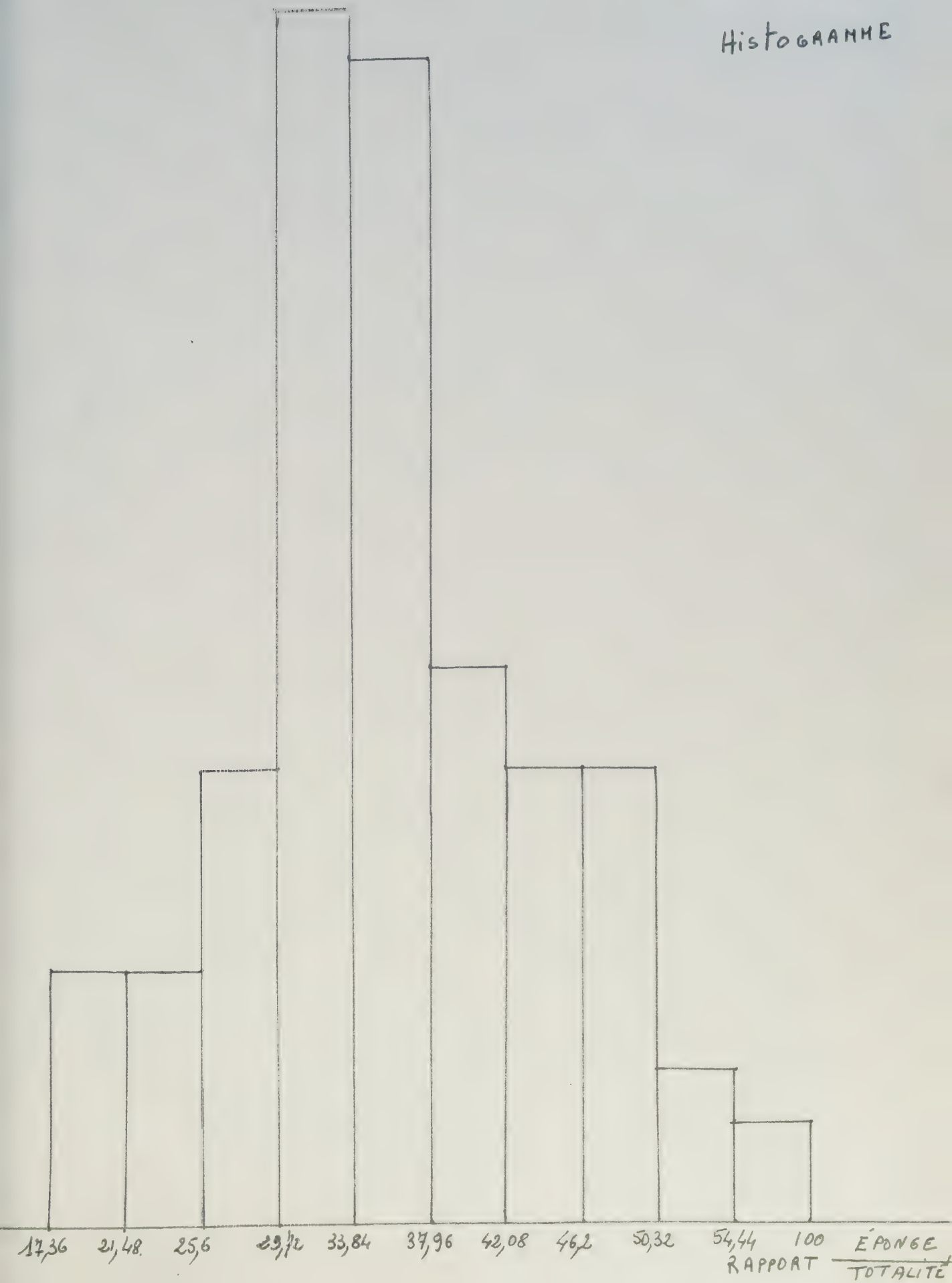
$$\chi^2 = \sum_{1}^{11} \frac{(Ri - Ti)^2}{Ti}$$

Avec $\nu = 6$; $\chi^2 = 8,109$

Ce lot peut donc être considéré comme distribué suivant une loi de Laplace-Gauss.

Nous avons aussi tracé un nuage de points, puis nous avons tracé la courbe de Laplace-Gauss, correspondant à cette distribution normale.

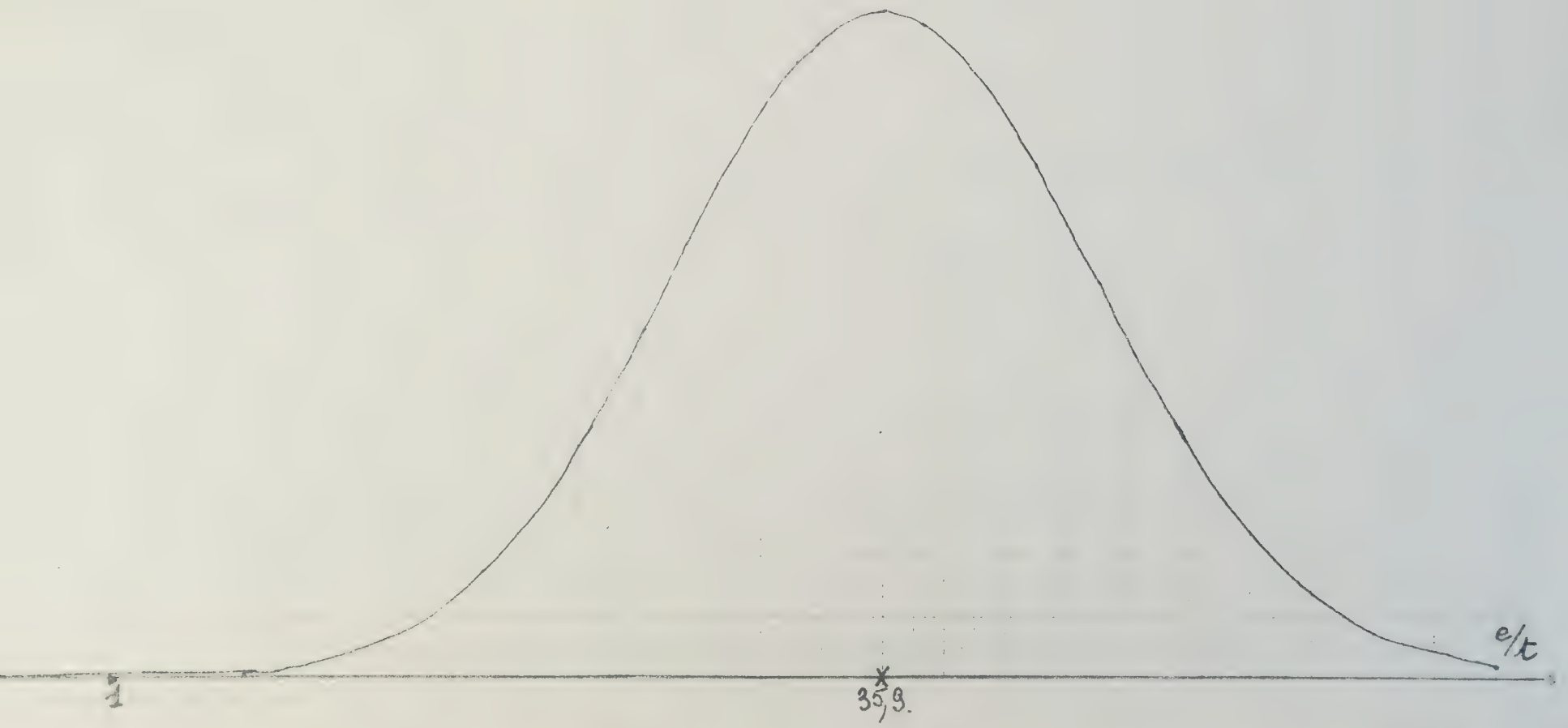
Histogramme

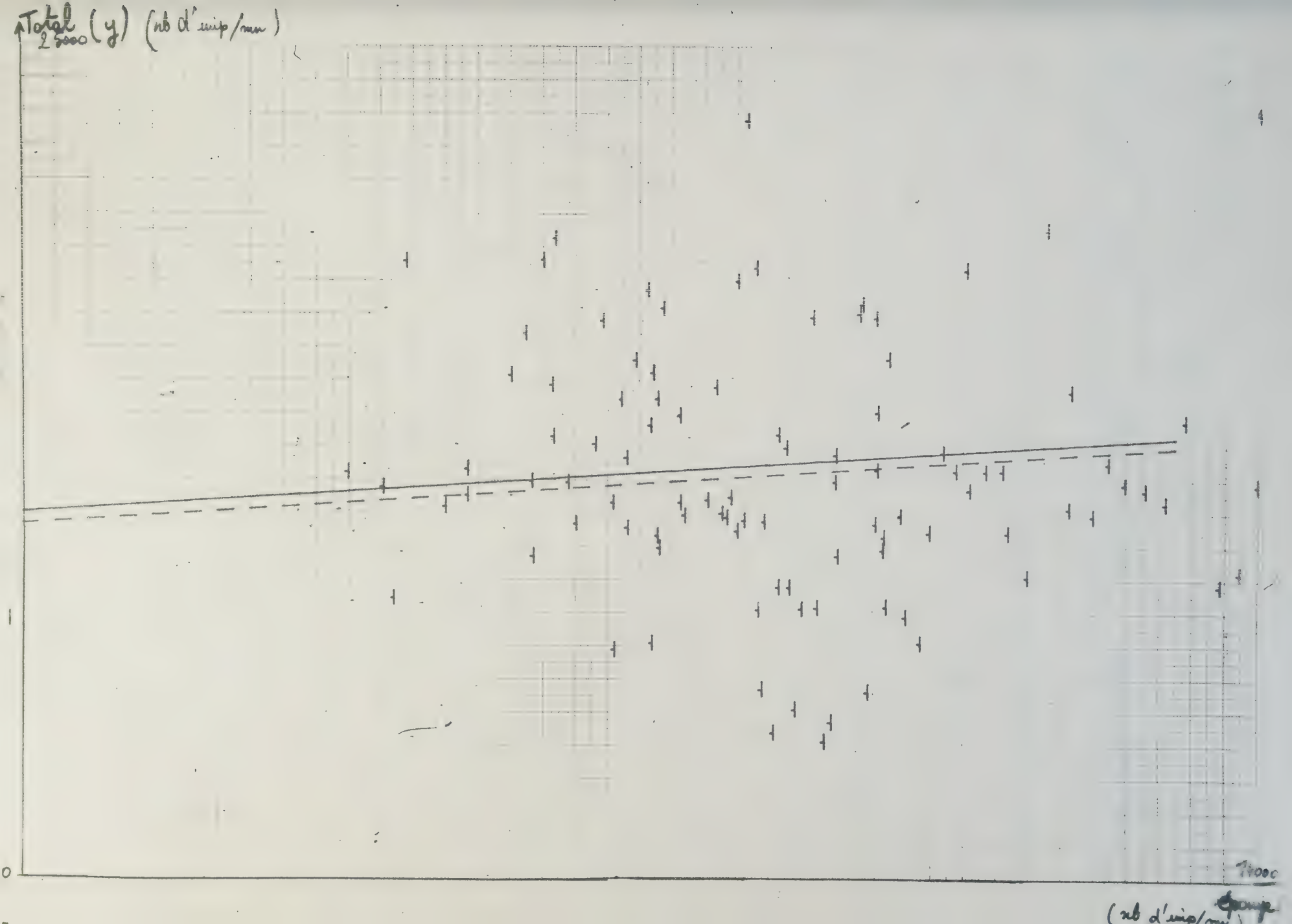


$E_{E/T}$ (e/t)

E/T est considéré comme une variable normale.

Tracé de la courbe de Gauss correspondant à ce lot




Total (y) (nb d'imp/mn)
25000
1
0
14000
Pompe
(nb d'imp/mn)

Etude du lot nº II

N = 100 Moyenne 34,63 % Ecart type : 8,47

SOUS CLASSE	LIMITE	EFFECTIF THEORIQUE	EFFECTIF EXPERIMENTAL
	0		
1		1,2224	0
	15,572		
2		2,7835	4
	19,807		
3		6,5591	3
	24,042		
4		12,098	4
	28,277		
5		17,466	31
	35,512		
6		19,742	21
	36,747		
7		17,466	10
	40,982		
8		12,098	9
	45,217		
9		6,5591	5
	49,452		
10		2,7835	4
	53,687		
11		1,2224	2
	100		

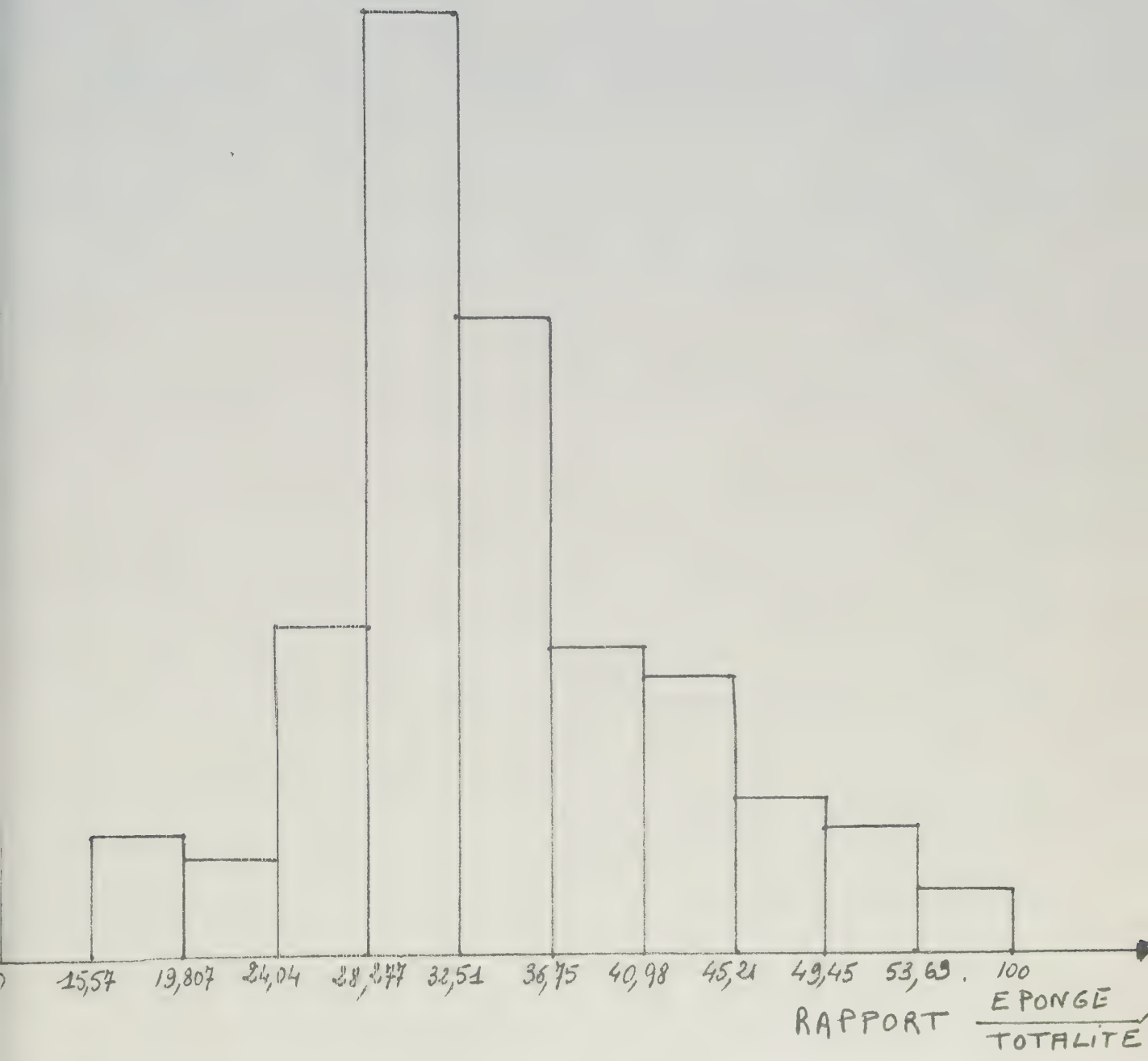
$\chi^2 = 16,033$ avec $\nu = 6$

La loi de distribution du rapport $\frac{E}{T}$ n'est donc pas normale pour les cent valeurs considérées.

nombre de malades

Lot n° II

Histogramme.



Etude du lot nº III -

N = 100 Moyenne 34,456 % variance 67,94 226

Ecart type 8,242

Coefficient de variation 0,239

SOUS CLASSE	LIMITE	EFFECTIF THEORIQUE	EFFECTIF EXPERIEMENTAL
	0		
1		1,2224	0
	15,909		
2		2,7835	2
	20,031		
3		6,5991	8
	24,152		
4		12,098	6
	28,273		
5		17,466	29
	32,395		
6		19 ,742	27
	36,516		
7		17,466	8
	40,638		
8		12,098	6
	44,759		
9		6,5991	6
	48,880		
10		2,7385	6
	53,002		
11		1,2224	2
	100		

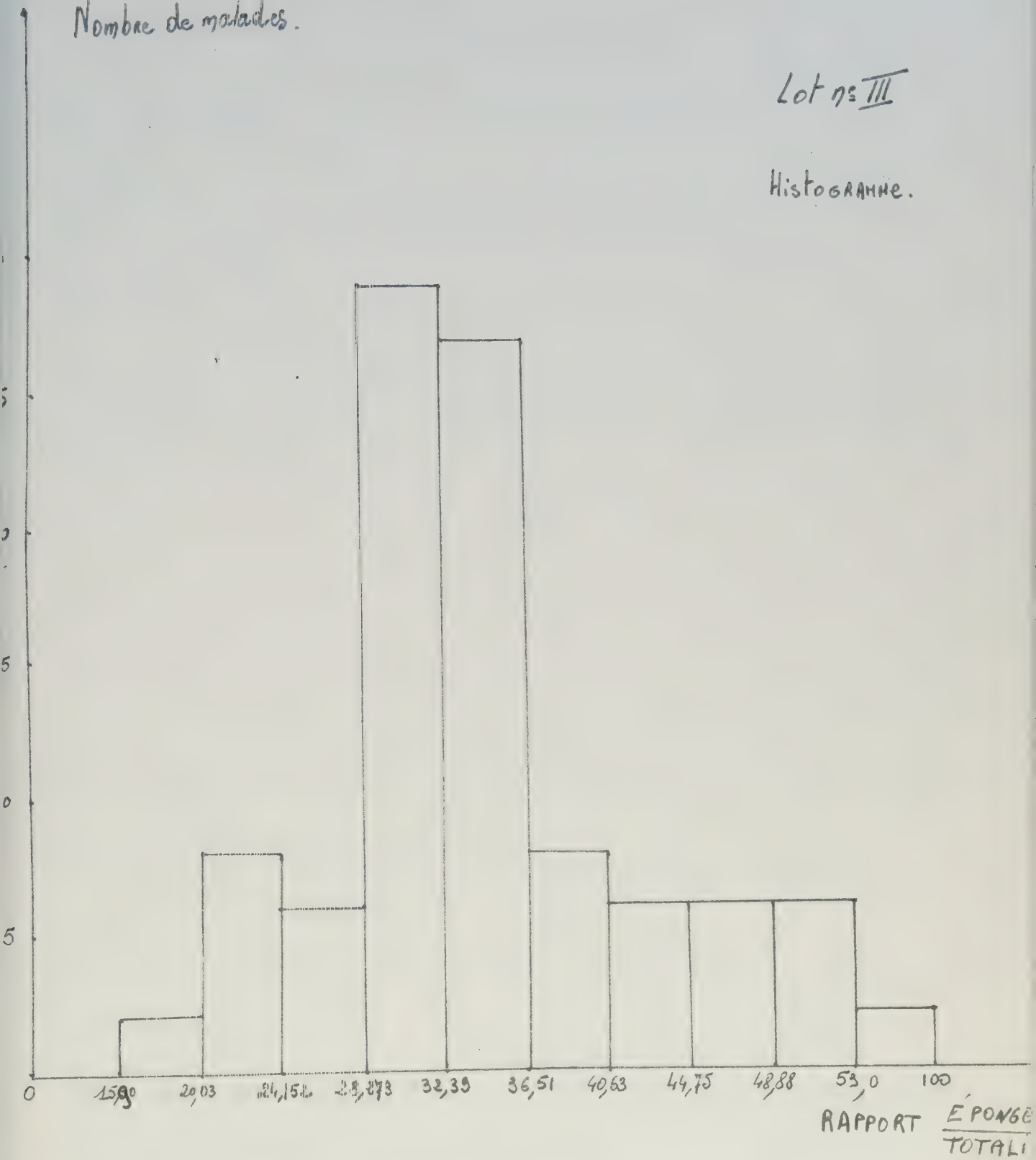
$\chi^2 = 22,709$ avec $\nu = 6$

D'où loi de distribution du rapport $\frac{E}{T}$ non normale pour les cent valeurs considérées.

Nombre de malades.

Lot n° III

Histogramme.



Etude du lot nº IV

N = 100

Moyenne = 34,419 %

Variance = 46,185

Ecart type = 6,795

Coefficient de variation 0,1974

SOUS CLASSE	LIMITES	EFFECTIF THEORIQUE	EFFECTIF EXPERIMENTAL
	0		
1		1,2224	0
	19,12		
2		2,7835	3
	22,52		
3		6,5591	5
	25,92		
4		12,098	14
	29,32		
5		17,466	15
	32,72		
6		19,742	27
	36,11		
7		17,466	14
	39,51		
8		12,098	13
	42,91		
9		6,5591	3
	46,31		
10		2,7835	3
	49,70		
11		1,2224	3
	100		

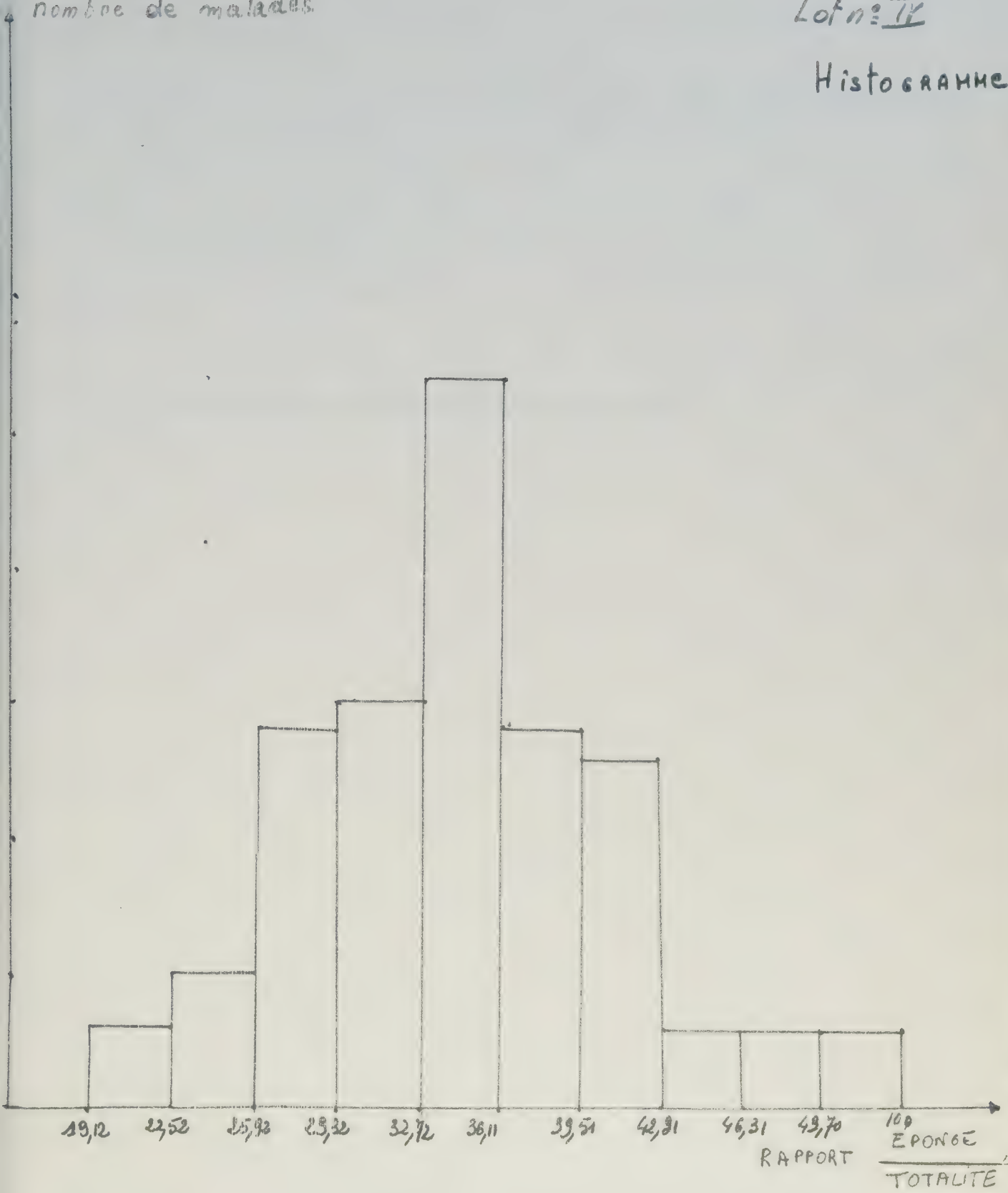
avec $\nu = 6$, nous trouvons $\chi^2 = 4,925$

Ce lot a donc une loi de distribution du rapport E/T, qui suit la loi de Laplace-Gauss. Nous voyons donc qu'il faut être très prudent.

nombre de malades

Lot n° IV

Histogramme



Supposons qu'au lieu d'étudier 1000 valeurs, nous en ayons étudié cent et justement les cent dernières valeurs dont nous parlons : lot nº IV.

A tort, on aurait pu écrire :

La loi du rapport des variables E/T est une variable normale.

Il aurait fallu écrire :

La loi du rapport des variables E/T est une loi normale pour les cent valeurs que nous avons étudiées.

Aussi, dans la première partie de notre travail, avons-nous précisé chaque fois : les valeurs sur lesquelles nous avons travaillé.

Etude du Lot nº V

N = 100

Moyenne = 44,28 %

Variance = 53,68

Ecart type = 7,32

Coefficient de variation 0,165

SOUS CLASSE	LIMITES	EFFECTIF REEL	EFFECTIF EXPERIMENTAL
	0		
1		1	1,2224
	27,79		
2		1	2,7835
	31,45		
3		4	6,5581
	35,12		
4		8	12,098
	38,78		
5		34	17,466
	42,44		
6		17	19,742
	46,4		
7		16	17,466
	49,77		
8		10	12,098
	53,43		
9		4	6,5591
	57,10		
10		3	2,7835
	60,76		
11		2	1,2224
	100		

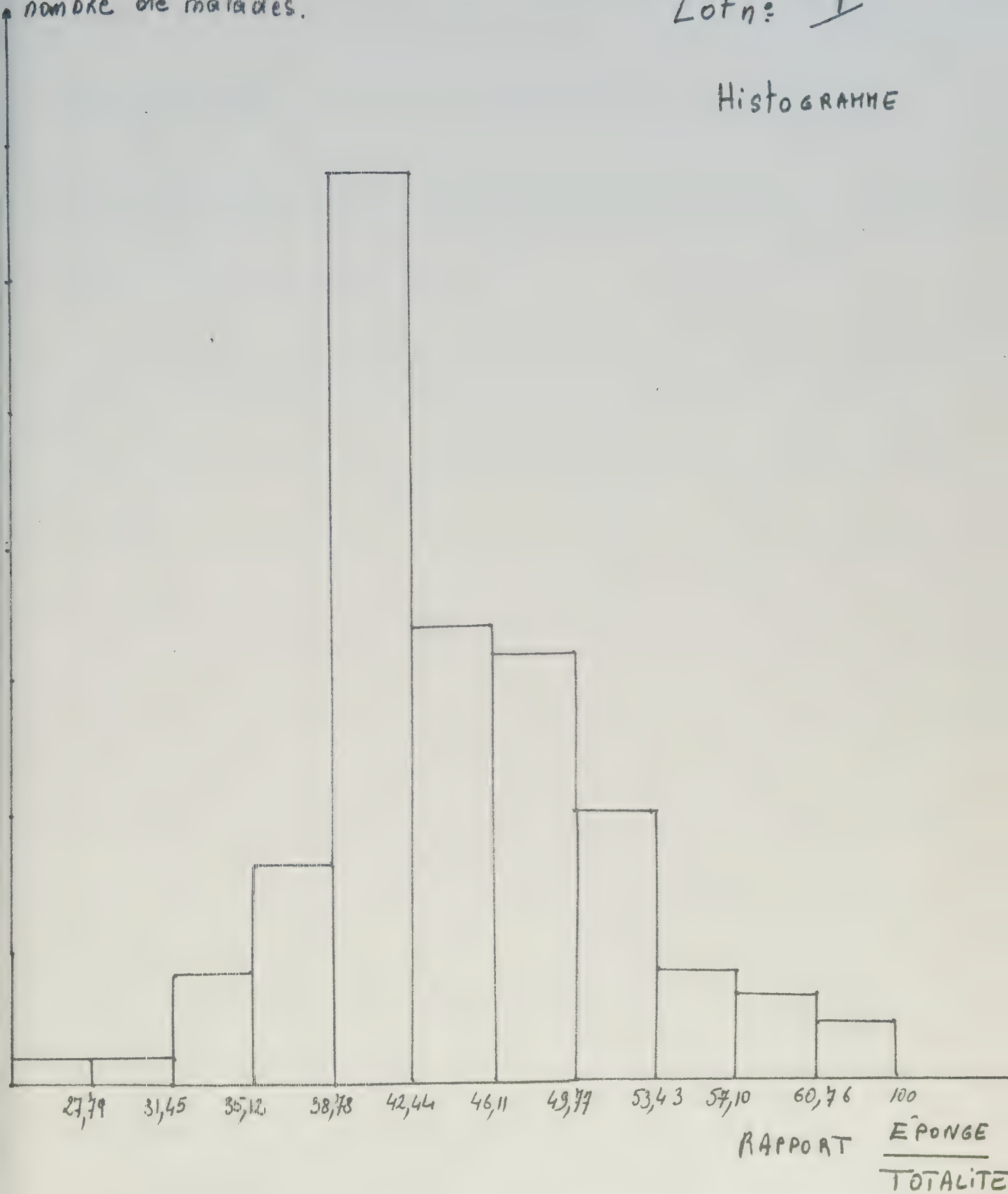
$\chi^2 = 20,110$ avec $\nu = 6$

Pour les valeurs du rapport $\frac{E}{T}$ de ce lot, nous pouvons dire que la loi de distribution n'est pas normale.

nombre de malades.

Lot n° V

Histogramme



Etude du lot nº VI

N = 100

Moyenne = 37,98 %

Ecart type = 7,76

Variance = 60,31

Coefficient de variation = 0,204

LIMITES DE CLASSES	SOUS-CLASSES	EFFECTIF THEORIQUE	EFFECTIF EXPERIM.
0			
	1	1,2224	1
20,507			
	2	2,7835	4
24,390			
	3	6,5591	1
28,273			
	4	12,098	12
32,156			
	5	17,466	22
36,039			
	6	19,742	29
39,922			
	7	17,466	15
43,805			
	8	12,098	5
47,688			
	9	6,5591	5
51,571			
	10	2,7835	2
55,454			
	11	1,2224	4
100			

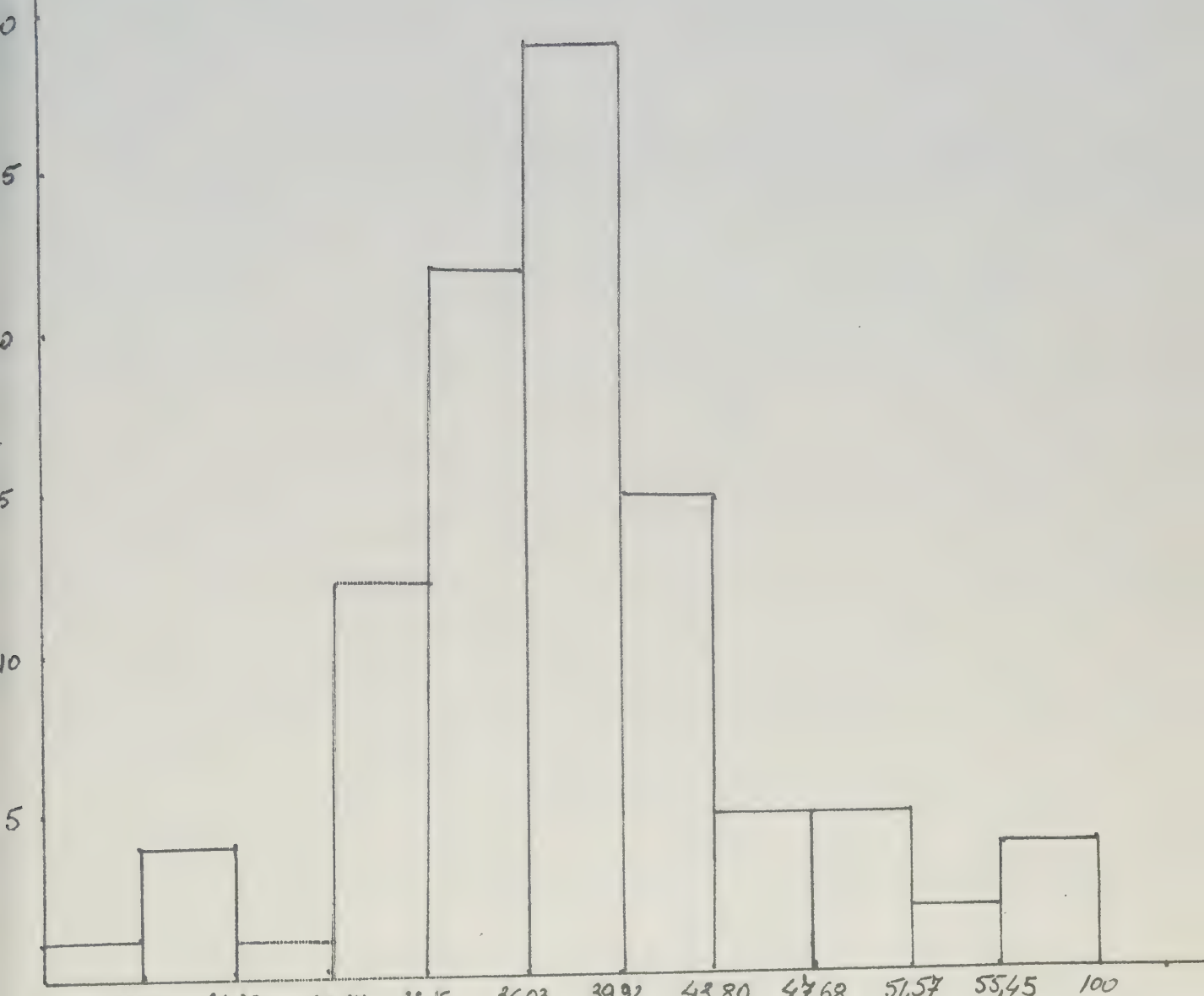
$\chi^2 = 11,964$ avec $\nu = 6$

La loi de distribution du rapport $\frac{E}{T}$ est normale, pour les cent valeurs considérées.

nombre de malades.

Lot n° VI

Histogramme.



RAPPORT ÉPON
TOTA

Etude du lot nº VII

N = 100

Moyenne = 34,921 %

Variance = 29,61

Ecart type = 5,44

Coefficient de variation : 0,155

SOUS CLASSES	LIMITES DE CLASSES	EFFECTIF THEORIQUE	EffECTIF EXPERIMENTAL
	0		
1		1,2224	1
	22,677		
2		2,7835	1
	25,398		
3		6,5591	3
	28,119		
4		12,098	12
	30,839		
5		17,466	27
	33,560		
6		19,742	25
	36,281		
7		17,466	14
	39,002		
8		12,098	8
	41,722		
9		6,5591	2
	44,443		
10		2,7835	2
	47,164		
11		1,2224	5
	100		

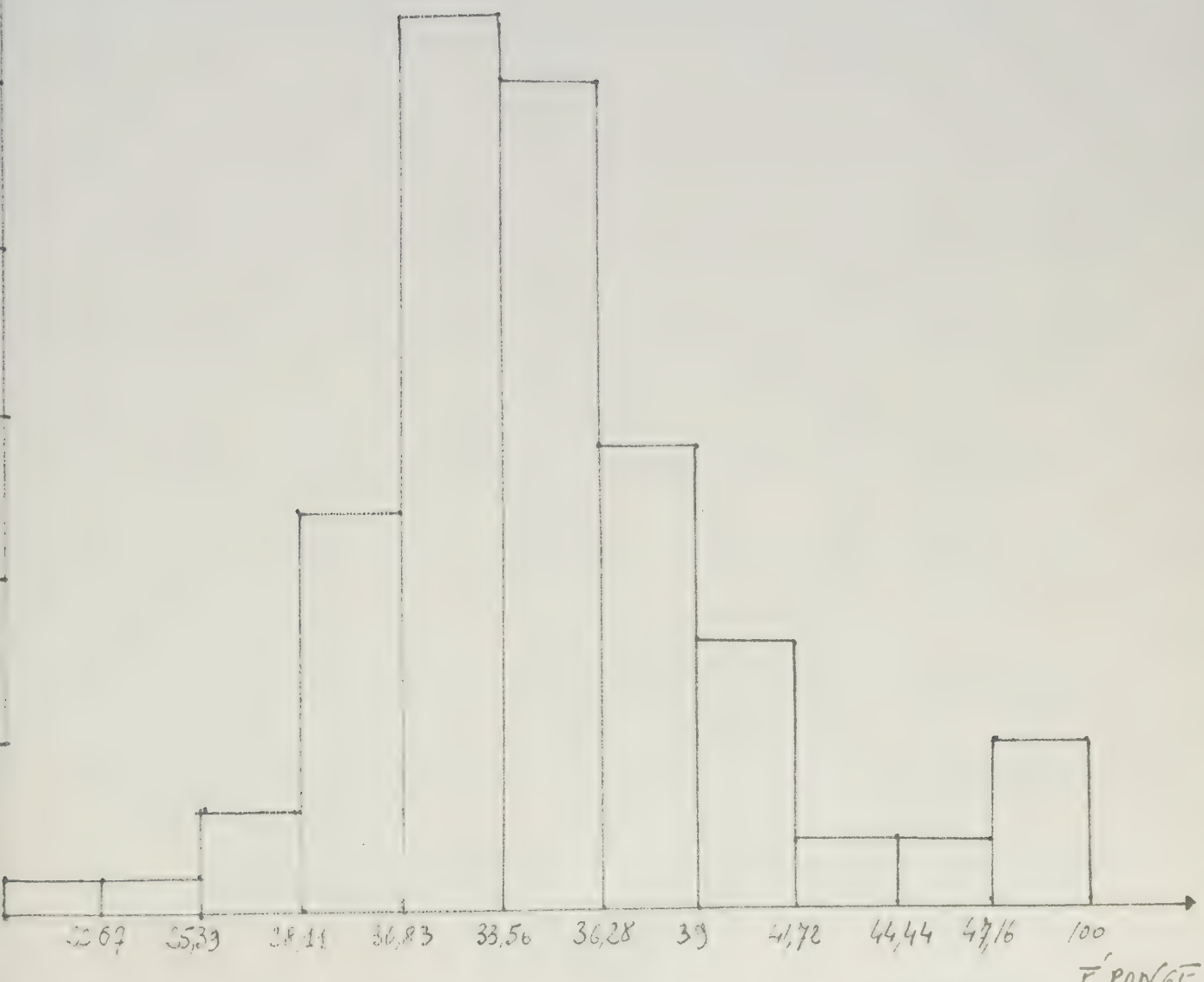
χ^2 = 11,842 avec ν = 6

La loi de distribution du rapport E/T est normale pour les cent valeurs considérées.

nombre de malades.

Lot nº VIII

Histogramme



RAPPORT $\frac{\text{ÉPONGE}}{\text{TOTALIT}}$

Etude du lot nº VIII

N =100

Moyenne = 44,37 %

Ecart type = 6,153

Variance 37,86

Coefficient de variation 0,138

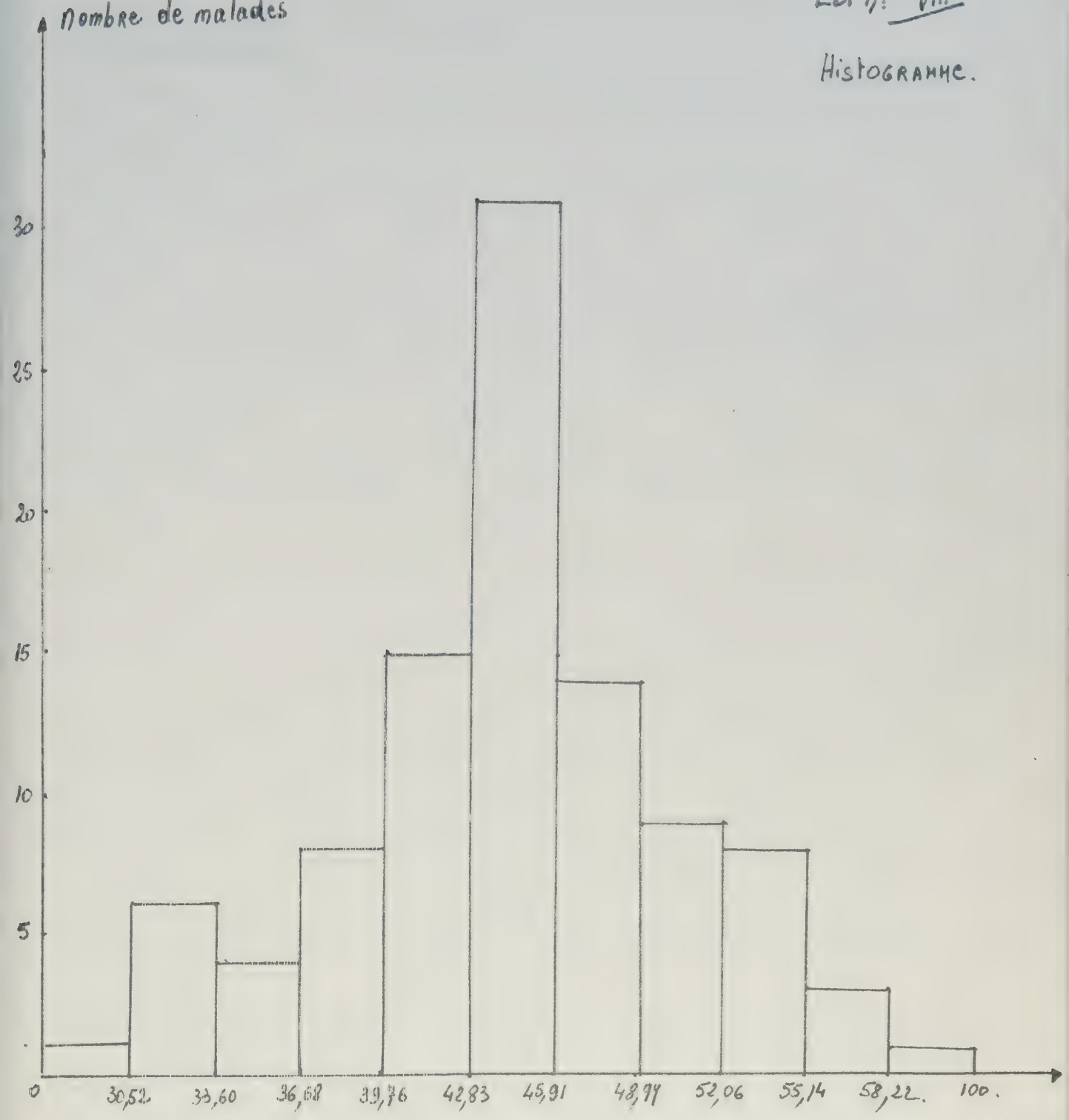
SOUS CLASSES	LIMITES	EFFECTIF THEORIQUE	EFFECTIF EXPERIMENTAL
	0		
1		1,2224	1
	30,529		
2		2,7835	6
	33,606		
3		6,5591	4
	36,683		
4		12,098	8
	39,760		
5		17,466	15
	42,837		
6		19,742	31
	45,914		
7		17,466	14
	48,991		
8		12,098	9
	52,068		
9		6,5591	8
	55,145		
10		2,7835	3
	58,222		
11		1,2224	1
	100		

$\chi^2 = 9,850$ avec $\nu = 6$

La loi de distribution du rapport des variables E/T peut donc être considérée comme normale, pour les cent valeurs considérées.

Lot n°: VIII

Histogramme.



Etude du lot nº IX

N = 100

Moyenne = 40,742 %

Ecart type = 7,46

Variance = 55,65

Coefficient de variation : 0,183

SOUS CLASSES	LIMITES	EFFECTIF THEORIQUE	EFFECTIF EXPERIMENTAL
	0		
1		1,2224	3
	23,956		
2		2,7835	2
	27,686		
3		6,5591	4
	31,416		
4		12,098	10
	35,146		
5		17,466	21
	38,876		
6		19,742	22
	42,607		
7		17,466	20
	46,337		
8		12,048	10
	50,067		
9		6,5591	4
	53,797		
10		2,7835	1
	57,527		
11		1,2224	3
	100		

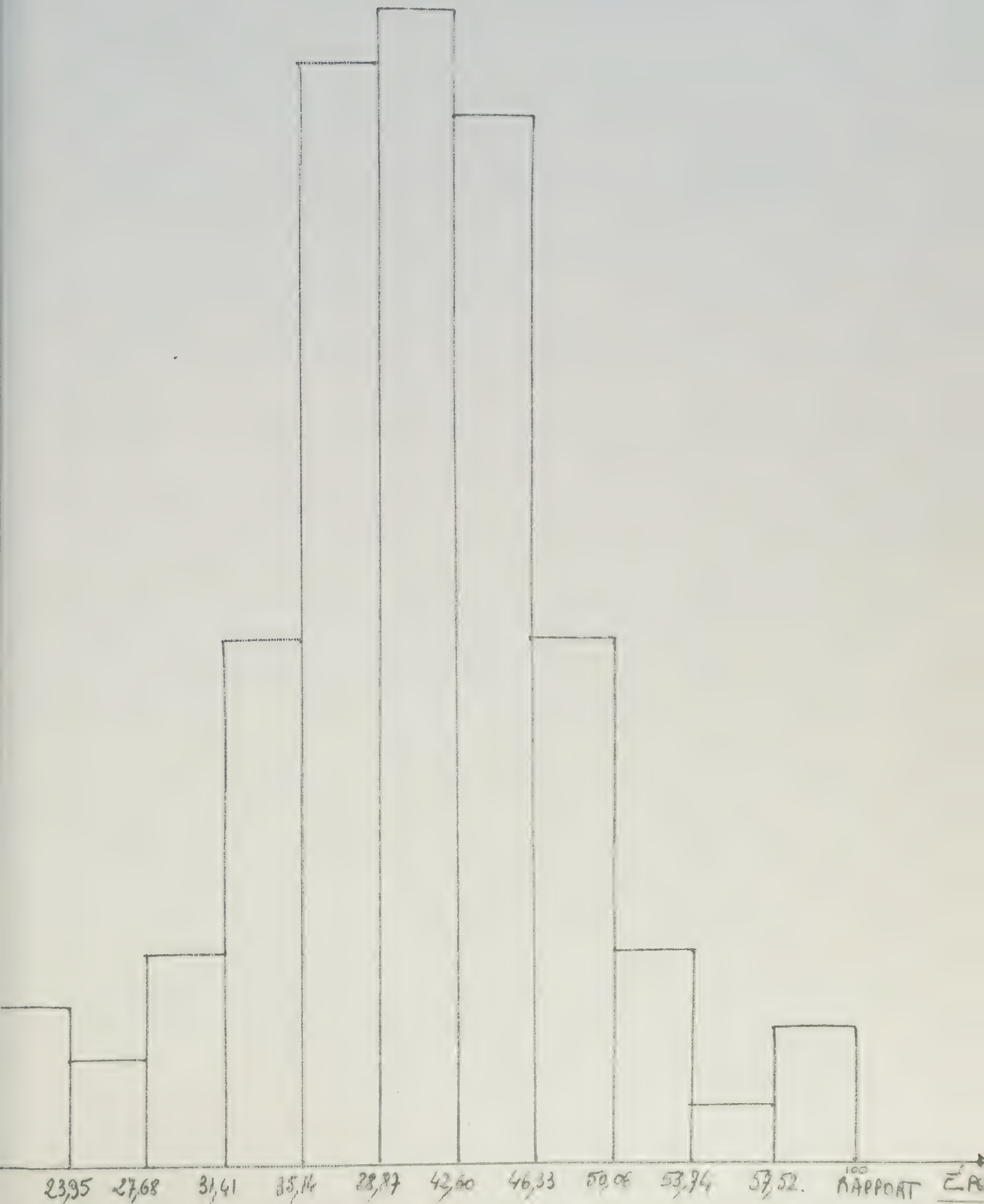
$\chi^2 = 2,923$ avec $\nu = 6$

La loi de distribution du rapport des variables E/T peut donc être considérée comme normale pour les œnt valeurs considérées.

nombre de malades

Lot n° IX

Histogramme



Etude du lot nº X

N = 100

Moyenne = 35,83

Ecart type = 8,33

Variance : 69,50

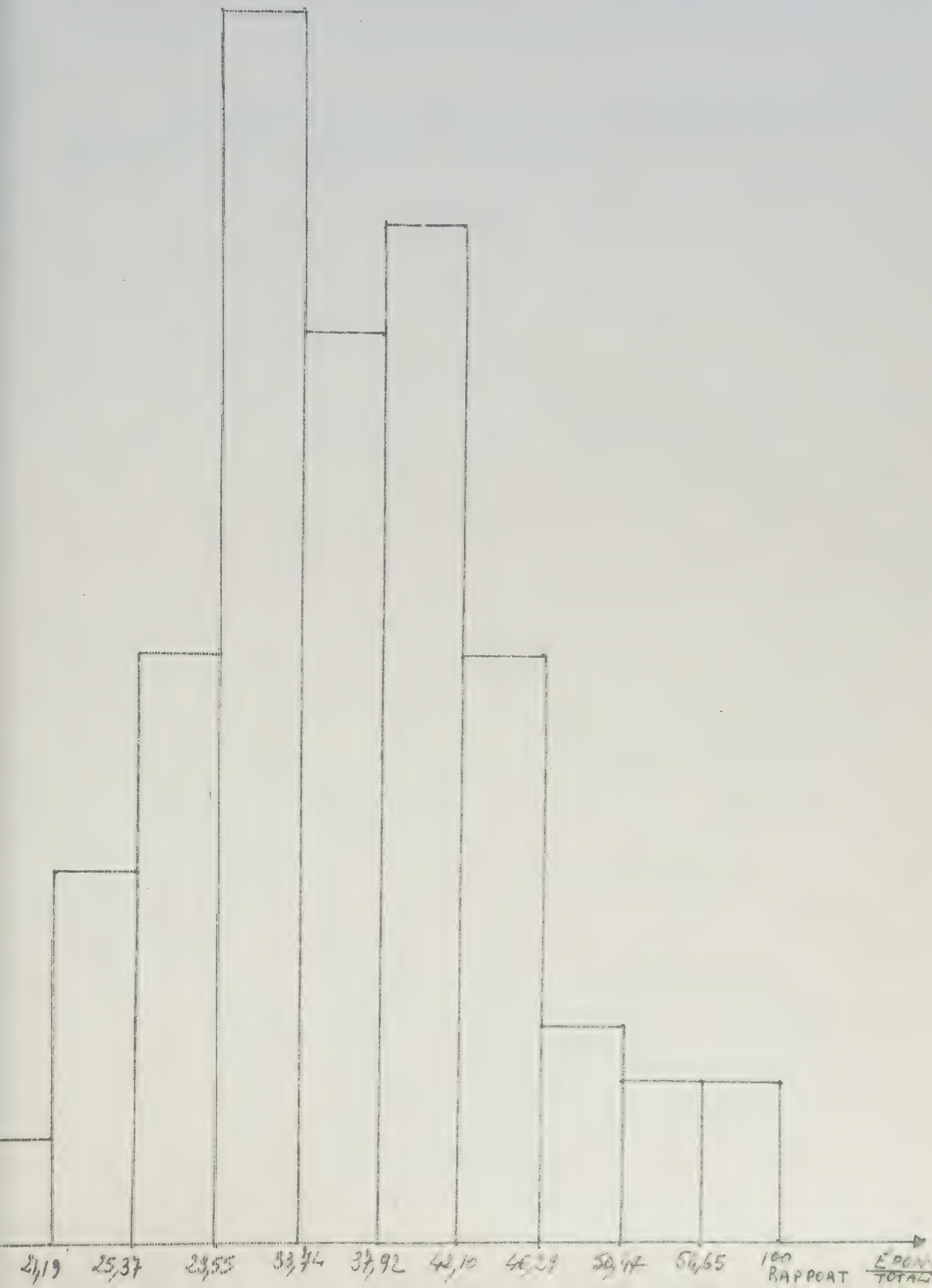
Coefficient de variation : 0,232

SOUS CLASSE	LIMITES	EFFECTIF THEORIQUE	EFFECTIF EXPERIMENTAL
	0		
1		1,2224	0
	17,007		
2		2,7835	2
	21,190		
3		6,5591	7
	25,374		
4		12,048	11
	29,557		
5		17,466	23
	33,741		
6		19,742	17
	37,924		
7		17,466	19
	42,108		
8		12,098	11
	46,291		
9		6,5591	4
	50,475		
10		2,7835	3
	54,658		
11		1,2224	3
	100		

ici : $\chi^2 = 2,931$ $\nu = 6$

Ici encore, la loi du rapport des variables E/T peut être considérée comme normale pour les cent valeurs considérées.

Histogramme.



Sur les dix lots, nous avons donc 3 lots au total, où la loi de distribution, du rapport E/T ne peut donc pas être considérée comme normale.

Nous avons ensuite étudié la population d'effectif N = 1000 variables E/T, en la considérant d'un seul bloc.

Nous avons, selon nos habitudes, calculé

sa moyenne : 37,7525
Son écart type : 8,3882
Sa variance : 70,3634
Son coefficient de variation : 0,2221

Ensuite, connaissant la moyenne et l'écart type, nous avons essayé d'ajuster la loi de distribution de notre population à une loi normale.

Nous avons divisé la population en 11 classes et nous avons calculé l'effectif théorique et nous l'avons comparé à l'effectif expérimental.

Ensuite, nous avons fait un calcul de χ^2, et nous avons dressé le tableau de contingence qui suit :

TABLEAU DE CONTINGENCE

SOUS CLASSES	LIMITES	EFFECTIF THEORIQUE	EFFECTIF EXPERIMENTAL
	0		
1		12,224	8
	18,878		
2		27,835	29
	23,072		
3		65,591	47
	27,267		
4		120,98	133
	31,461		
5		174,66	219
	35,655		
6		197,42	182
	39,849		
7		174,66	164
	44,043		
8		120,98	109
	48,237		
9		65,591	61
	52,432		
10		27,835	27
	56,626		
11		12,224	21
	100		

Pour le calcul du χ^2

$\chi^2 = 28,928$ avec $\nu = 10$

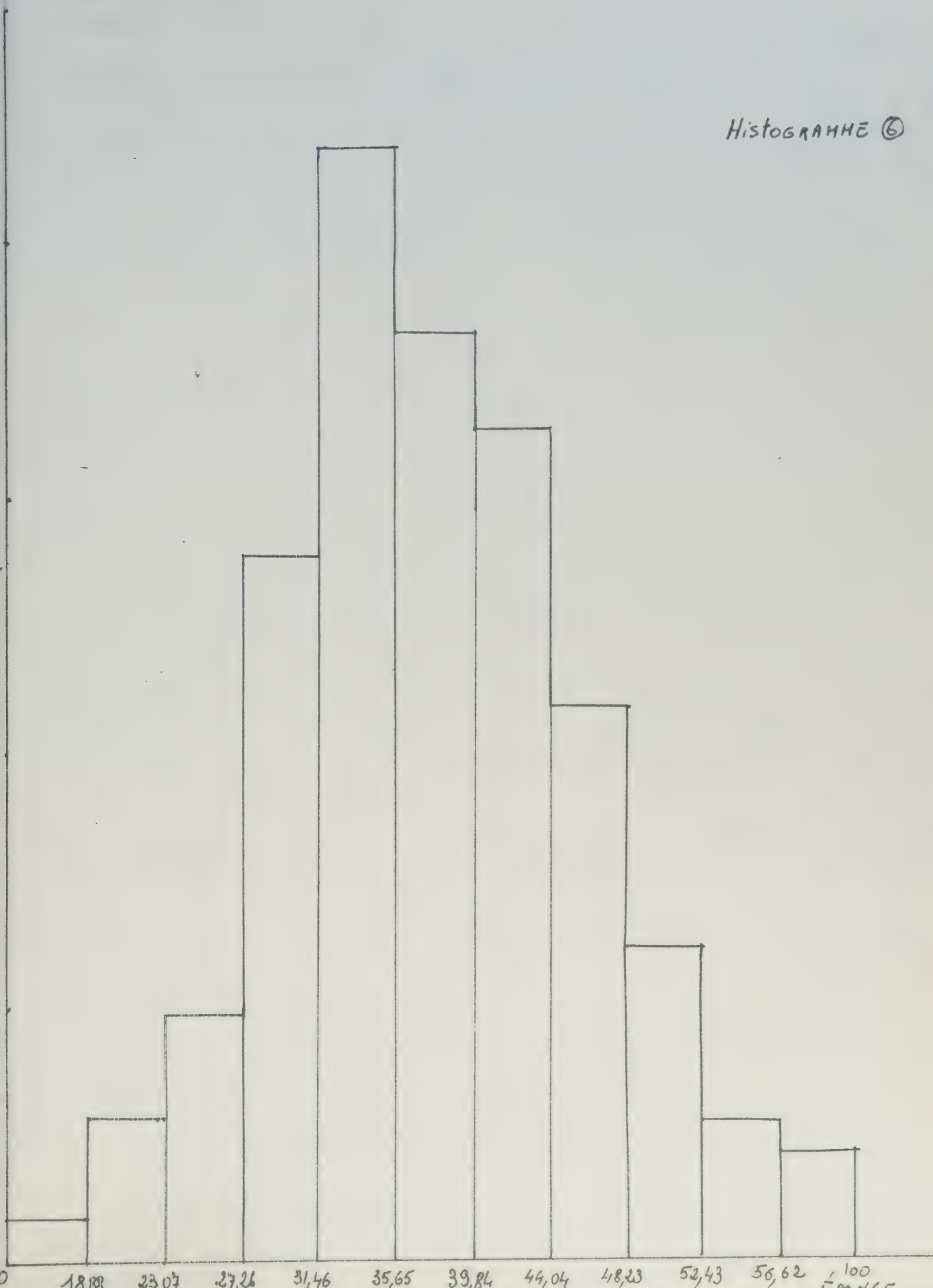
Ce χ^2 exprime, que la loi de distribution de la variable E/T, n'est pas une loi normale.

Donc, pour résumer :

- La variable E n'est pas distribuée normalement
  (Le χ^2 vaut $\simeq$ 680 avec $\nu = 10$).
- La variable T n'est pas distribuée normalement.
  (Le χ^2 vaut $\simeq$ 159 avec $\nu = 10$)
- La variable E/T n'est pas distribuée normalement
  (Le χ^2 vaut $\simeq$ 29 avec $\nu = 10$).

nb de malades.

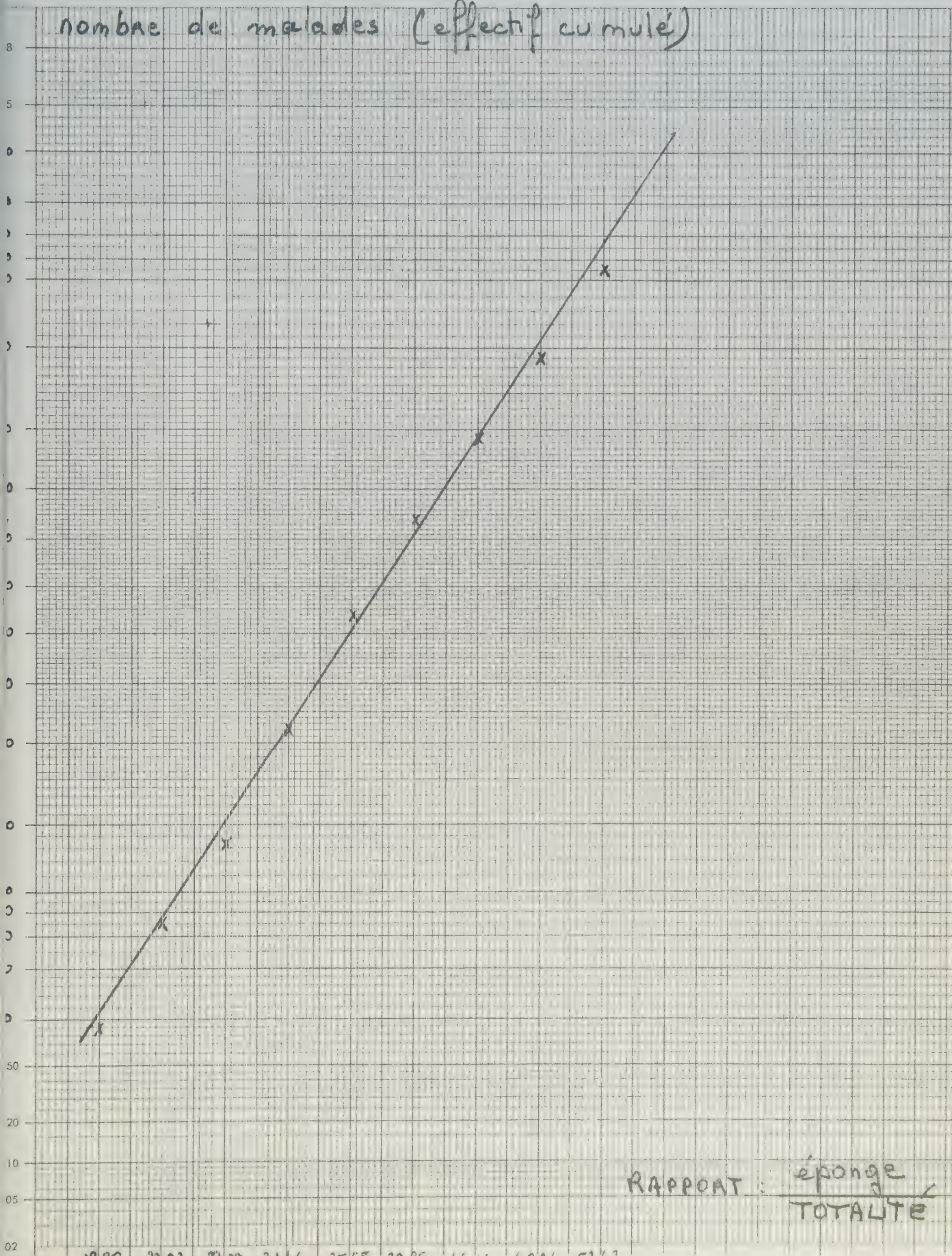
Histogramme ⑥



)DUIT :

:actéristique:

nombre de malades (effectif cumulé)



RAPPORT : éponge / TOTALITÉ

Bien que la variable E/T ne soit pas distribuée normalement, nous voyons qu'elle est celle qui se rapproche le plus d'une variable normale.

Ainsi, il apparaît que le rapport de deux variables distribuées non normalement, tend plus vers la loi normale que les variables elles-même. (Il suffit d'ailleurs de consulter les histogrammes et les tracés des droites de Henry, pour s'en apercevoir.)

Nous avons comparé les χ^2 de la population étudiée en bloc, d'une part, et étudiée par classes d'effectif N = 100 d'autre part.

Pour la population prise en bloc :

Avec $\nu = 10$ nous avons $\chi^2 = 28,93$; (N = 1000)

Pour la population divisée en 10 classes :

Nous avons calculé la moyenne des χ^2 des 10 classes, nous avons trouvé :

$\chi^2 = 11,13$ avec $\nu = 6$.

Bien que le nombre de degrés de liberté ne soit pas le même ($\nu = 6$ et $\nu = 10$) nous voyons naturellement que lorsque l'effectif de la population augmente, le χ^2 augmente.

Nous voyons aussi qu'il faut être très prudent avant de déclarer une population normale ou non, ou alors préciser "pour les variables que j'ai étudiées".

V - Etude de la liaison entre les variables aléatoires "Eponge" et "Totalité"

Appelons "x" les valeurs que peut prendre la variable aléatoire "éponge" et "y", les valeurs que peut prendre la variable aléatoire "totalité".

Nous avons calculé les expressions :

$$\sum_{0}^{1000} x = 7965\ 563 \rightarrow \bar{x} = 7\ 965,563$$

$$\sum_{0}^{1000} x^2 = 72\ 310\ 879\ 300$$

$$\sum_{0}^{1000} y = 20\ 931\ 589 \rightarrow \bar{y} = 20\ 931,589$$

$$\sum_{0}^{1000} y^2 = 441\ 364\ 985 \times 10^3$$

En appareillant les valeurs de "x" et "y" pour un même malade, nous avons calculé l'expression $\sum_{0}^{1000} x.y.$ Nous avons trouvé : 167 465 922 000

La covariance (X,Y) est donc égale à 734 031.

Ainsi, nous avons pu calculer le coefficient de corrélation ρ (X,Y)

$$\rho\ (X,Y) = \frac{cov\ (X,Y)}{\sigma X \times \sigma Y} = 0,137.$$

Ce coefficient est significatif. Il montre que "E" et "T" sont deux variables très peu liées entre elles.

Nous avons aussi calculé l'équation de la droite de régression de Y sur X.

$$Y' = 20\ 271,71 + 0,082\ X$$

ainsi que celle de la droite de régression de X sur Y

$$X' = 3\ 214,01 + 0,22\ Y$$

et la somme des carrés des écarts à la droite de régression de Y sur X:

$$\sum_{0}^{1000} (Y - Y')^2 = 3\ 172\ 756\ 900$$

ainsi que la somme des carrés des écarts à la droite de régression de X sur Y.

$$\sum_{0}^{1000} (X - X')^2 = 8\ 694\ 057\ 700$$

Variance de Y, X connue: 3 179 115,1.

Variance de X, Y connue : 8 711 480,7

Nous avons représenté le phénomène par un nuage de points en prenant seulement 100 malades (voir la figure présentée auparavant).

Cette étude nous montre donc bien, que X et Y ne sont pas liées.

VI - Etude de la liaison entre les variables aléatoires "Eponge" et le rapport $\frac{\text{éponge}}{\text{totalité}}$

Appelons "x" la valeur que peut prendre la variable éponge et $\frac{x}{y}$ les valeurs que peut prendre la variable $\frac{\text{éponge}}{\text{totalité}}$.

Nous avons calculé les expressions :

$$\sum_{0}^{1000} x = 7\ 965\ 563 \rightarrow \bar{x} = 7\ 965{,}563$$

$$\sum_{0}^{1000} x^2 = 72\ 310\ 879\ 300$$

$$\sum_{0}^{1000} x/y = 37\ 752{,}5 \rightarrow \left[\frac{\bar{x}}{y}\right] = 37{,}752$$

$$\sum_{0}^{1000} \frac{x^2}{y^2} = 1\ 495\ 614{,}75$$

$$\sum x.y = 314\ 900\ 134$$

Nous avons ainsi calculé la covariance $(X/\underline{\frac{X}{Y}})$, nous avons trouvé : 14 180,217.

Le coefficient de correlation $\rho(X, \underline{\frac{X}{Y}})$ a donc, comme valeur : 0,567, valeur qui ne nous étonne pas d'ailleurs, vues les variables que nous avons prises.

Nous avons aussi calculé les équations des droites de regression de Y sur X, avec : $Y = \frac{x}{y}$

$$Y' = 0{,}016\ X + 25{,}004$$

et de X sur Y

$$X' = 201{,}52\ Y + 357{,}37$$

somme des carrés des écarts à la droite de regression de Y sur X

$$\sum_{0}^{1000} (Y - Y')^2 = 47\ 670{,}14$$

somme des carrés des écarts à la droite de regression de X sur Y.

$$\sum_{0}^{1000} (X - Y')^2 = 6\ 002\ 973\ 930$$

Variance de Y, X connu : 47,76

Variance de X, Y connu : 6 015 003,9.

Applications

Avec des variables non normales, nous nous sommes demandés dans quelle mesure pouvait-on appliquer la formule de Geary.

Nous avons calculé le rapport :

Moyenne des variables "Totalité"

Ecart type des variables "Totalité"

Nous avons trouvé : 11,6 ($\gg$ 3)

Appelons :

μY : la moyenne des valeurs des variables "Totalité".

μX : la moyenne des valeurs des variables "Eponge".

Nous avons :

$\mu Y = 20\ 931,58$

$\mu X = 7\ 965,56$

σY : l'écart type correspondant aux variables "Totalité".

σX : l'écart type correspondant aux variables "Eponge".

$\sigma Y = 1798,21$

$\sigma X = 2976,69$

D'après Geary :

$$\frac{x}{y} = \frac{\mu X}{\mu Y} + u\sqrt{\frac{(\sigma X)^2}{(\mu Y)^2} + \left(\frac{\sigma Y^2}{\mu Y^2} - 2\rho\frac{\sigma X \cdot Y}{\mu X \mu Y} + \frac{\sigma X^2}{\mu X^2}\right)}$$

$\rho = 0,137$

$$\frac{x}{y} = 0,3805 + u \times 0,1414$$

$\frac{\text{Eponge}}{\text{Totalité}}$ D'après l'approximation de Geary, la moyenne des valeurs des rapports est 0,3805. Expérimentalement nous trouvons : 0,3775.

L'écart type, d'après l'approximation de Geary est : 0,1414. L'écart réel est 0,0838.

Que faut-il penser de la différence entre ces deux écarts types ?

Utilisons le test Snédécor :

$V_1 = S_1^2$; $V_2 = S_2^2$ $\qquad V_1 > V_2$

$$F = \frac{(0,1414)^2}{(0,0838)^2} = 2,84 \ (\gg 1)$$

Ceci nous montre que nous ne pouvons pas considérer que l'approximation de Geary s'applique lorsque les variables ne sont plus des variables normales.

C O N C L U S I O N S

Nous avons montré l'influence de l'effectif d'une population sur la valeur du χ^2 et l'importance qu'il y a de préciser le nombre de variables sur lesquelles on travaille.

Nous avons donné deux exemples d'application de l'approximation de Geary:

l'un où les variables étaient normales,

l'autre où les variables n'étaient pas normales.

Nous avons montré qu'avec des variables normales, l'approximation de Geary donnait de bons résultats, alors qu'elle ne s'applique plus lorsque les variables ne sont pas normales,. ce qui montre combien l'approximation est restrictive et l'approximation peu robuste.

Nous avons pu calculer les paramètres dont dépend la loi du rapport de deux variables normales.

Enfin, il nous est apparu, en faisant l'étude des χ^2, que le rapport de deux variables non normales, tend plus vers la loi normale que les variables elles-mêmes.

B I B L I O G R A P H I E

1 - BAILAR (J.C) et EDERER (F) - Significance factors for the ratio of a Poisson variable to its expectation - Biometrics 1964, 20, p.639-643

2 - BASS (J) - Eléments de calcul de probabilités théoriques et appliquées Masson édit. Paris.

3 - BENETT (G) - Probability inequalities for the sum of independent random variables - JASA.1962, 57, p.33-45

4 - BLOCH (D) - A note on the estimation of the location parameter of the Cauchy distribution - JASA.1966, 61, p. 852-855

5 - BOERNSTEDT (G.W) et GOLDBERGER (A.S) - On the exact covariance of products of random variables - JASA. 1969, 64, p. 1439-1442

6 - BUEHLER (R.J) - Confidence intervals for the product of two binomial parameters - J.A.S.A. 1957, 52 , p. 482-493

7 - CHIANG (C.L) - On the expectation of the reciprocal of a random variable. The American Statistician - 1966, 20, nº 4, p. 28

8 - DAVID (H.A) et NORMAN (J) - Exact distribution of the sum of independent identically distributed discret random variables - J.A.S.A - 1965 - 60 p. 837-842

9 - FELLER (W) - An introduction to probability theory and its applications. Vol. I et II - John Willey and sohns, 2ème édit. 1966

10 - FLEISS (J.L) - A note on the expectation of the reciprocal of a random variable - The American Statistician - 1966 - 20, nº 1, 25

11 - GOODMAN (L.A) - The variance of the product of k random variables - J.A.S.A - 1962 - 57, p. 54-60

12 - GOVINDARAJULV (Z) - The reciprocal of the decapited negative binomial variable - J.A.S.A. - 1962 - 57, p. 906-913

13 - GUPTA (S) et PILLAI (K.C.S) et STECK (G.P) - On the distribution of linear functions and ratios of linear functions of ordered correlated normal random variables with emphasis on range - BIOMETRIKA - 1964, 51, 143-151

14 - GURLAND (J) - Inequalities of expectation of random variables derived by monotonicity or convexity - THE AMERICAN STATISTICIAN - 1968, 22, nº 2 p. 26-27

15 - GURLAND (J) - An inequality satisfied by the expectation of the reciprocal of a random variable - THE AMERICAN STATISTICIAN - 1967 - 21, nº 2, p.24-25

16 - HALD (A) - Statistical Theory with engineering - Applications - WILEY publications in Statistics - 6ème édition - 1965

17 - HELLER (R) - Manuel de Statistiques Biologiques - GAUTHIERS VILLARS édit. Paris - 1958

18 - HINLEY (D.V) - On the ratio of two correlated normal random variables - BIOMETRIKA - 1969 - 56, p. 635-639

19 - HOYLE (M.H) - Letter to the Editor - THE AMERICAL STATISTICIAN - 1967, 21, nº 3, 41

20 - JOHNSON (N.L) - Letter to the Editor the American statistician 1967, 21, nº 3 - 40

21 - KENDALL (M.G) et STUART (A) - The advanced theory of statistics - 1 vol. GRIFFIN édit. Londres, 1958.

22 - KOOP (J.C) - An exercise in ratio estimatics - THE AMERICAN STATISTICIAN 1968 - 22 - nº 1, p.29-30

23 - KOOP (J.C) - Letter to the éditor - THE AMERICAN STATISTICIAN, 1967, 21, nº 3, 39

24 - LORD (E) - The use of the range in place of the standard deviation in the "t" test - BIOMETRIKA - 1947, 34, p. 41-47

25 - MARSAGLIA (G) - Ratios of normal variables and ratio of sums of uniform variables - J.A.S.A. - 1965, 60, p. 193-204

26 - MULLEN (K) - A note on the ratio of two independent random variables. THE AMERICAN STATISTICIAN - 1967, 21, nº 3 - p.30-31

27 - MURTHY (V.N) and PILLAI (K.C.S) - A note on the expectation of the reciprocal and square root of a random variable - THE AMERICAN STATISTICIAN 1966 - 20, nº 5 p. 30

28 - PRESS (S.J) - The ratio distribution - J.A.S.A - 1969 - 64, 242-252

29 - RAJ (D) - A note on the covariance of the ratio estimation - J.A.S.A. 1964 - 59 - p. 895-898

30 - RAO (R.S) - Comparison of four ratio type estimates under a model - J.A.S.A 1969 - 64, p. 574-580

31 - RIDER (P.R.) - Distribution of product and quotient of maximum values in samples from a power function population. - J.A.S.A. - 1964, 59, p. 877-880

32 - RIDER (P.R.) - Expected values and standard deviation of the reciprocal of a variable from a decapited negative binomial distribution - J.A.S.A. 1962 - 57, p. 439-445

33 - ROTHENBERG (T.J) FISCHER (F) et TILANUS (C.B) - A note on estimation from a Cauchy sample - J.A.S.A. - 1964, 59, p. 460-463

34 - ROUAIS - APPORTS DES TESTS IN VITRO utilisant les indicateurs nucléaires dans l'exploration du métabolisme thyroïdien - Thèse BORDEAUX - 1969 nº 589

35 - SCLOVE (S), SIMONS (G) et VAN RYZIN (J) - Further remarks on the expectation of the reciprocal of a positive random variable. THE AMERICAN STATISTICIAN - 1967, 21, nº 4, p. 33-34

36 - SERFLING (R.J) - A note on the covariance of a random variable and its reciprocal - The American Statistician - 1967, 21, nº 4, p. 33

37 - SHEPS (M.C) - Characteristics of a ratio used to estimate failure rates : occurences per person year of exposure - BIOMETRICS, 1966,22, p. 310-321

38 - SMITH (H.F) - Approximate formulare for bias and variance of ratios THE AMERICAN STATISTICIAN - 1969 - 23, nº 4, p. 29

39 - SORIN (B) et THIONET (P) - Lois de probabilités de Bessel - REVUE DE STATISTIQUES APPLIQUEES - 1960, 16, nº 4, p. 65-72